

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 337.008 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712A-1_COPY_1_1695

Perfect score: 1695
Sequence: 1 gtcacgaatctgcagcaaga.....cagtactgtcattgcatta 1695

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2195239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1980.DAT:*

2: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1981.DAT:*

3: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1982.DAT:*

4: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1983.DAT:*

5: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1984.DAT:*

6: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1985.DAT:*

7: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1986.DAT:*

8: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1987.DAT:*

9: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1988.DAT:*

10: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1989.DAT:*

11: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1990.DAT:*

12: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1991.DAT:*

13: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1992.DAT:*

14: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1993.DAT:*

15: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1994.DAT:*

16: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1995.DAT:*

17: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1996.DAT:*

18: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1997.DAT:*

19: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1998.DAT:*

20: /SID22/gcgdata/geneseq/geneqseq-emb1/NA1999.DAT:*

21: /SID22/gcgdata/geneseq/geneqseq-emb1/NA2000.DAT:*

22: /SID22/gcgdata/geneseq/geneqseq-emb1/NA2001A.DAT:*

23: /SID22/gcgdata/geneseq/geneqseq-emb1/NA2001B.DAT:*

24: /SID22/gcgdata/geneseq/geneqseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1695	100.0	2024	21	AAA07697 Human collectin en
2	1695	100.0	2628	22	AAH43036 Nucleotide sequenc
3	1693.4	99.9	2262	24	ABA97932 Human scavenger re
4	1693.4	99.9	2641	22	AAC66903 Human EXMAD-14 cod
5	1681.4	99.2	2929	24	ABQ92072 Human polynucleoti
6	1681.4	99.2	2930	20	AAV55746 Human secreted pro
7	1672.8	98.7	2181	22	AAI60628 Human polynucleoti
8	1661.8	98.0	2318	22	AAI58842 Human polynucleoti
9	1477.6	87.2	2005	24	ABL95574 Human angiogenesis

10	1477.6	87.2	2005	24	ABL88085 Human PRO7223 CDNA
11	1359	80.2	2637	22	AAH43037 Nucleotide sequenc
12	941	55.5	2256	22	AAH43054 Nucleotide sequenc
13	883.8	52.1	1521	23	AAH71133 DNA encoding novel
14	248.2	14.6	3685	20	AAH27858 Human CSR3 protein
15	248.2	14.6	3810	20	AAH27856 DNA encoding novel
16	169.4	10.0	493	23	AAH71131 DNA encoding novel
17	150.2	8.9	873	23	AAH71130 DNA encoding novel
18	150.2	8.9	1062	23	AAH71134 DNA encoding novel
19	146.8	8.7	1329	23	AAH71134 DNA encoding novel
20	143	8.4	5041	22	ABA17153 Human nervous syst
21	135.6	8.0	1877	20	AAH27857 Human CSR2 protein
22	133.6	7.9	1707	22	AAH44989 CDNA encoding nove
23	127.2	7.5	986	22	AAI89468 Human polynucleoti
24	121.6	7.2	756	14	AAH43034 Collagen-like poly
25	121.6	7.2	756	17	AAH16768 Collagen-like poly
26	121.4	7.2	5676	15	AAH64556 Human collagen (Ty
27	121.4	7.2	8284	22	AAH28526 Human breast cance
28	121.4	7.2	8284	22	AAH26553 Human breast cance
29	121.4	7.2	8284	22	AAH26600 Murine Col5a3 cDNA
30	118.4	7.0	6109	23	AAH78667 Collagen-like poly
31	118	7.0	756	14	AAH43032 Collagen-like poly
32	118	7.0	756	17	AAH16766 Human benign prost
33	117	6.9	9287	24	ABK64501 Bovine alpha1(I) c
34	115.8	6.8	4748	22	AAH06573 Mouse ischaemic co
35	115	6.8	4270	24	ABH9885 Human cDNA differe
36	114.8	6.8	4816	24	ABK84041 DNA encoding novel
37	114.8	6.8	5058	23	AAH69026 Human EST-derived
38	114.8	6.8	6358	22	AAH98286 DNA encoding novel
39	114.8	6.8	6691	23	AAH79806 DNA encoding novel
40	114	6.7	5060	24	ABK92176 Prostate cancer-as
41	113.8	6.7	506	17	AAH12198 pJG4-5-CDK-BP CDNA
42	113.6	6.7	3181	19	AAV59358 Nucleotide sequenc
43	113.4	6.7	4821	23	AAH68666 DNA encoding novel
44	113.4	6.7	6200	23	AAH78668 Human COL5A3 CDNA
45	113.4	6.7	6368	24	ABN59647 Novel human coding

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX
AC AAA07697;
XX
DT 20-JUN-2000 (first entry)
XX
DE Human collectin encoding DNA.
XX
KW Collectin; human; antibacterial; antiviral; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 55..1698
FT /tag= a
FT /product= "collectin"
FT misc_feature 1..738
FT /tag= b
FT /note= "fragment specifically claimed in claim 8"
FT misc_feature 55..738
FT /tag= c
FT /note= "fragment specifically claimed in claim 8"
FT misc_feature 79..738
FT /tag= d
FT /note= "fragment specifically claimed in claim 8"
FT misc_feature 325..738
FT /tag= e
FT /note= "fragment specifically claimed in claim 8"
FT misc_feature 358..738
FT /tag= f

FT misc_feature /note= "fragment specifically claimed in claim 8"
 FT 670...1695
 FT /tag= g
 FT /note= "fragment specifically claimed in claim 5"
 FT 685...738
 FT /tag= h
 FT /note= "fragment specifically claimed in claim 7"
 FT 730...738
 FT /tag= i
 FT /note= "fragment specifically claimed in claim 7"
 FT 739...1695
 FT /tag= j
 FT /note= "fragment specifically claimed in claim 6"
 FT 1696...2024
 FT /tag= k
 FT /note= "fragment specifically claimed in claim 9"
 FT
 XX WO200011161-A1.
 PN
 XX
 XX 02-MAR-2000.
 PD
 XX
 XX 24-AUG-1999; 99WO-JP04552.
 PF
 XX
 XX 24-AUG-1998; 98JP-0237611.
 PR
 XX
 XX (FUSO) FUSO PHARM IND LTD.
 PA
 XX
 XX Wakamiya N;
 PI
 XX
 XX WPI: 2000-224696/19.
 DR
 DR P-PSDB; AAY77985.
 DR
 XX
 XX New collectin of human origin having antibacterial and antiviral
 PT activity, and gene encoding it useful for production of transgenic
 PT animals and of antibodies for screening potential drug molecules
 PT
 XX
 XX Claim 5-9; Page 87-90; 106pp; Japanese.
 PS
 XX
 XX The invention relates to polynucleotides encoding a new collectin of
 CC human origin. The collectin can be used as an antibacterial and antiviral
 CC agent and for screening potential drug molecules. The new collectin can
 CC be produced by standard recombinant methodology. The present sequence
 CC represents a DNA encoding the human collectin.
 XX
 XX Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;
 SQ
 Query Match 100.0%; Score 1695; DB 21; Length 2024;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCACGAATCTGCAGCAAGATACAGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
 DB 1 GTCACGAATCTGCAGCAAGATACAGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
 QY 61 TCTCATATGTGTCATCATGAACTCAACACCTGAACCTGACCGGTGTCAGCAGAGAG 120
 DB 61 TCTCATATGTGTCATCATGAACTCAACACCTGAACCTGACCGGTGTCAGCAGAGAG 120
 QY 121 AACCTCATCAGAACTGTCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAAATC 180
 DB 121 AACCTCATCAGAACTGTCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAAATC 180
 QY 181 AAGAAGCACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACACGGATTGG 240
 DB 181 AAGAAGCACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACACGGATTGG 240
 QY 241 CTGAAGGAGAAAGTCAGAGCTTGCAGACGCTGGCTGCGCAACAACTCTCGGTTGGCCAAA 300
 DB 241 CTGAAGGAGAAAGTCAGAGCTTGCAGACGCTGGCTGCGCAACAACTCTCGGTTGGCCAAA 300
 QY 301 GCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGTCAAGT 360
 DB 301 GCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGTCAAGT 360

QY 361 GAGAACATCACCACCTATCTCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCGAGGACTTA 420
 DB 361 GAGAACATCACCACCTATCTCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCGAGGACTTA 420
 QY 421 CACAAAGATGCGAGAGAAATGAGACAGCCATCAAGTTCAACCACTGAGGAGAACCTTCCAG 480
 DB 421 CACAAAGATGCGAGAGAAATGAGACAGCCATCAAGTTCAACCACTGAGGAGAACCTTCCAG 480
 QY 481 CTCTTTTGACACCGATATGTGACATCATAGCAATATAGTATACACAGCCACACCTG 540
 DB 481 CTCTTTTGACACCGATATGTGACATCATAGCAATATAGTATACACAGCCACACCTG 540
 QY 541 CGGACCTGACACGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCTTACCAAA 600
 DB 541 CGGACCTGACACGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCTTACCAAA 600
 QY 601 CACACAGATGATCTGACCTCTTGAATATATACCTGGCCAACTCCGTTTGGATTCTGTT 660
 DB 601 CACACAGATGATCTGACCTCTTGAATATATACCTGGCCAACTCCGTTTGGATTCTGTT 660
 QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCAACTTA 720
 DB 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCAACTTA 720
 QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT 780
 DB 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT 780
 QY 781 TTTTACAATACTACAGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCCAG 840
 DB 781 TTTTACAATACTACAGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCCAG 840
 QY 841 GGACCCCTCGGGCCCACTGSCAACAGGACAGAAAGGAGAGAGGGAGGCTGACCA 900
 DB 841 GGACCCCTCGGGCCCACTGSCAACAGGACAGAAAGGAGAGAGGGAGGCTGACCA 900
 QY 901 CCTGGCCCTCGGGTGGAGAGAGGCCCCAATTTGGACCACTGGTCCCGCCGAGAGCGTGGC 960
 DB 901 CCTGGCCCTCGGGTGGAGAGAGGCCCCAATTTGGACCACTGGTCCCGCCGAGAGCGTGGC 960
 QY 961 GGCAGAGATCTAAAGGCTCCAGGGCCCGCAAGGCTCCGCTGGTTCCTCGGAGAGCCC 1020
 DB 961 GGCAGAGATCTAAAGGCTCCAGGGCCCGCAAGGCTCCGCTGGTTCCTCGGAGAGCCC 1020
 QY 1021 GGCCCTCAGGGCCCGAGTGGGACCCAGGCCCCCGGGCCCGCCAGGCAAGAGGACTC 1080
 DB 1021 GGCCCTCAGGGCCCGAGTGGGACCCAGGCCCCCGGGCCCGCCAGGCAAGAGGACTC 1080
 QY 1081 CCGGCCCCCTCAGGGCCCTCTCGCTTCCAGGGACTTTCAGGGCACCGTTGGGAGCCTGGG 1140
 DB 1081 CCGGCCCCCTCAGGGCCCTCTCGCTTCCAGGGACTTTCAGGGCACCGTTGGGAGCCTGGG 1140
 QY 1141 GTGCTTGAGACTTGGGGACTGCCAGGCTTGCCTGGGGTACCAGGATGCCAGGCCCAAG 1200
 DB 1141 GTGCTTGAGACTTGGGGACTGCCAGGCTTGCCTGGGGTACCAGGATGCCAGGCCCAAG 1200
 QY 1201 GGCCCCCGGGCCCTCTCGGCCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAAGATGAG 1260
 DB 1201 GGCCCCCGGGCCCTCTCGGCCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAAGATGAG 1260
 QY 1261 CCAACCCCGGCGCCGAGGACCAATGGCTCCCGCCCTCACTGGAAGAACTTCACAGACAAA 1320
 DB 1261 CCAACCCCGGCGCCGAGGACCAATGGCTCCCGCCCTCACTGGAAGAACTTCACAGACAAA 1320
 QY 1321 TGCCTACTATTTTTCAGTTCAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGAC 1380
 DB 1321 TGCCTACTATTTTTCAGTTCAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGAC 1380
 QY 1381 AAGTCTTCACATCTTCTTTTTCATAAACTAGAGAGGAAACAGCAATGGATAAAAAACAG 1440
 DB 1381 AAGTCTTCACATCTTCTTTTTCATAAACTAGAGAGGAAACAGCAATGGATAAAAAACAG 1440

QY 1441 ATGGTAGGAGAGAGAGCCACTGGATCGGCTCAGAGCTCAGAGCTGAAATGAATGG 1500
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1501 AAGTGGCTGGATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGATAC 1560
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1561 AAGTGGCTGGATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGATAC 1560
QY 1561 TGGGCTCATGGCCATGGGCGAGGAGAACTGTGCTGGGTTGATTTATGCTGGGAGTGG 1620
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1621 AAGTGGCTGGATGGGCGAGGAGAACTGTGCTGGGTTGATTTATGCTGGGAGTGG 1620
QY 1621 AACGATTTCCAAATGGAAGACGTCATTAACCTTTTCATTTGCGAAAAGACAGGAGACAGTA 1680
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1681 CTGCTCATCTGCATTA 1695
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1681 CTGCTCATCTGCATTA 1695
RESULT 2
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX
AC AAH43036;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRC1-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 74..2302
FT /*tag= a
FT /product= "scavenger receptor"
XX
PN WO200159107-A1.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-JP00874.
XX
PR 14-FEB-2000; 2000JP-0035155.
PR 10-OCT-2000; 2000JP-0309068.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
DR WPI; 2001-497076/54.
DR P-PSDB; AAG63346.
XX
PT New Scavenger receptor proteins SRC1-PI with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation -
XX
PS Claim 2; Page 79-84; 118pp; Japanese.
XX
CC The present sequence encodes a human scavenger receptor, designated
CC SRC1-PI. The SRC1-PI polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.

XX SQ Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;
Query Match 100.0%; Score 1695; DB 22; Length 2628;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTAGGATCTCGACGAAGATACCGCGTGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 605 GTCTAGGATCTCGACGAAGATACCGCGTGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 664
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 TCTCATATGTGGTTCATCATGAACCTCAACAACCTGAACTGACCCAGGTGCAGCAGAG 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 665 TCTCATATGTGGTTCATCATGAACCTCAACAACCTGAACTGACCCAGGTGCAGCAGAG 724
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 AACCTCATCACGAATCTGCAGGGTGTGTGGATGACACAAGCCAGGCTATCCAGCAATC 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 725 AACCTCATCACGAATCTGCAGGGTGTGTGGATGACACAAGCCAGGCTATCCAGCAATC 784
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 AAGAACGACTTTCAAAATCTGCAGCAGGTTTTTTCTCAAGCCAAAGACACGGATTGG 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 785 AAGAACGACTTTCAAAATCTGCAGCAGGTTTTTTCTCAAGCCAAAGACACGGATTGG 844
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGGTGCCAACAACCTCTGCGTTGGCCAA 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 845 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGGTGCCAACAACCTCTGCGTTGGCCAA 904
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 301 GCCAACACGACACCTCGGAGGATATGAACGCCAGCTCAACTCATTCACAGTCAGATG 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 905 GCCAACACGACACCTCGGAGGATATGAACGCCAGCTCAACTCATTCACAGTCAGATG 964
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 361 GAGAACATCACCACTATCTCTCAAGCCAAAGACGAGCAACCTGAAAGACCTGCAGGACTTA 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 965 GAGAACATCACCACTATCTCTCAAGCCAAAGACGAGCAACCTGAAAGACCTGCAGGACTTA 1024
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 421 CACAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACCAACACTGTGAGGAACGCTTCCAG 480
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1025 CACAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACCAACACTGTGAGGAACGCTTCCAG 1084
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 481 CTCCTTGAGACGATATTTGTGAACATCATTAGCAATATCAGTTACACACGCCACCACTTG 540
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1085 CTCCTTGAGACGATATTTGTGAACATCATTAGCAATATCAGTTACACACGCCACCACTTG 1144
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 600
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1145 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 1204
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 601 CACACAGATGATCTGACCTCTTGAATATATACCTGGCCAAACATCCGTTTGATTCGTT 660
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1205 CACACAGATGATCTGACCTCTTGAATATATACCTGGCCAAACATCCGTTTGATTCGTT 1264
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTGCAGGTTAGACACTGAAGTAGCAACTTA 720
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1265 TCTCTCAGGATGCAACAAGATTTGATGAGTGCAGGTTAGACACTGAAGTAGCAACTTA 1324
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 721 TCAGTGATTTGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTGCAGCTCATCAAGAT 780
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1325 TCAGTGATTTGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTGCAGCTCATCAAGAT 1384
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 781 TTTTACAATACTACAGGTCCACCGGCCCCAGGGTTCAGAGGTGCAGAGAGATCCAG 840
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1385 TTTTACAATACTACAGGTCCACCGGCCCCAGGGTTCAGAGGTGCAGAGAGATCCAG 1444
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 841 GGACCCCTGGCCCCACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCTTGACCA 900
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1445 GGACCCCTGGCCCCACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCTTGACCA 1504
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 901 CTGGCCCTTCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGC 960
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1505 CTGGCCCTTCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGC 1564
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 961 GGCAAGGATCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTGGGAAGGCC 1020
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 1565 GCGAAGGATCTAAAGGCTCCCAAGGCCCCCAAGAGGCTCCCGTGGTTCCTCCCTGGGAAGCC 1624
 QY 1021 GGCCCTCAGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGACTC 1080
 Db 1625 GGCCCTCAGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGACTC 1684
 QY 1081 CCGGCCCCCAGGCCCCCTCTGGCTTCCAGGACTTCCAGGACCCGTTGGGGAGCCCTGGG 1140
 Db 1685 CCGGCCCCCAGGCCCCCTCTGGCTTCCAGGACTTCCAGGACCCGTTGGGGAGCCCTGGG 1744
 QY 1141 GTGCTGGGACTCGGAGACTGCCAGCTTCCCTGGGTTCCAGGACCCGTTGGGGAGCCCTGGG 1200
 Db 1745 GTGCTGGGACTCGGAGACTGCCAGCTTCCCTGGGTTCCAGGACCCGTTGGGGAGCCCTGGG 1804
 QY 1201 GGCCCCCCCCGCTCTCTGGCCCATCAGGAGCGTGGTGGCCCTGGCCCTGCAGAAATGAG 1260
 Db 1805 GGCCCCCCCCGCTCTCTGGCCCATCAGGAGCGTGGTGGCCCTGGCCCTGCAGAAATGAG 1864
 QY 1261 CCAACCCCGGACCGGAGGACAATGGCTGCGCCCTCAGTGGAGAACTTTCACAGACAAA 1320
 Db 1865 CCAACCCCGGACCGGAGGACAATGGCTGCGCCCTCAGTGGAGAACTTTCACAGACAAA 1924
 QY 1321 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCCAAGCTTTTCTGTGAAGAC 1380
 Db 1925 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCCAAGCTTTTCTGTGAAGAC 1984
 QY 1381 AAGCTTTCACATCTGCTTTTATATAACACTAGAGAGGACAGCAATGGATATAAAAAACAG 1440
 Db 1985 AAGCTTTCACATCTGCTTTTATATAACACTAGAGAGGACAGCAATGGATATAAAAAACAG 2044
 QY 1441 ATGGTAGGAGAGAGAGCCACTGGATGGCCCTCACAGACTCAGAGCGTGAATGAATGG 1500
 Db 2045 ATGGTAGGAGAGAGAGCCACTGGATGGCCCTCACAGACTCAGAGCGTGAATGAATGG 2104
 QY 1501 AAGTGGCTGGATGGGACATCTCAGACTACAAAAATTTGAAAGCTGGAGCGCGGATAAC 1560
 Db 2105 AAGTGGCTGGATGGGACATCTCAGACTACAAAAATTTGAAAGCTGGAGCGCGGATAAC 2164
 QY 1561 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGCGAGTGG 1620
 Db 2165 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGCGAGTGG 2224
 QY 1621 AAGCATTTCAATGTGAAGACGTCATAATCTTCATTTGGGAAAAAGACAGGAGACAGTA 1680
 Db 2225 AAGCATTTCAATGTGAAGACGTCATAATCTTCATTTGGGAAAAAGACAGGAGACAGTA 2284
 QY 1681 CTGTCATCTGCATTA 1695
 Db 2285 CTGTCATCTGCATTA 2299

RESULT 3
 ABA97932
 ID ABA97932 standard; cdna; 2262 BP.
 XX AC ABA97932;
 XX DT
 XX 25-APR-2002 (first entry)
 XX Human scavenger receptor-like protein encoding cdna SEQ ID NO 1.
 XX Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
 KW rheumatoid arthritis; gene; ss.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 FT CDS 33..2262
 FT /*tag= a
 FT /product= "scavenger receptor-like protein"
 XX JP2001340089-A.

PD 11-DEC-2001.
 XX 08-DEC-2000; 2000JP-0375066.
 XX 27-MAR-2000; 2000JP-0090772.
 XX (SHIO) SHIONOGI & CO LTD.
 XX WPI: 2002-144965/19.
 DR P-PSDB; ABB08642.
 XX New scavenger receptor-like protein for diagnosis, prevention and
 PT treatment of autoimmune disease, such as rheumatoid arthritis -
 PS Claim 3; Fig 1; 38pp; Japanese.
 CC The invention relates to a human scavenger receptor-like protein. The
 CC protein is useful as a target molecule for diagnosis, prevention and
 CC treatment of autoimmune diseases such as rheumatoid arthritis.
 XX Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
 Query Match 99.9%; Score 1693.4; DB 24; Length 2262;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1694; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCACGAATCTGCAGCAAGATACACGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
 Db 565 GTCACGAATCTGCAGCAAGATACACGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 624
 QY 61 TCTCATTAATGTGTCTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGTCAGCAGAGG 120
 Db 625 TCTCATTAATGTGTCTCATCATGAACCTCAACAACCTGACCCAGGTGTCAGCAGAGG 684
 QY 121 AACCTCATCAGAAATCTGACGGTCTGTGGATGACACAGCCAGCTATCCAGCCAATC 180
 Db 685 AACCTCATCAGAAATCTGACGGTCTGTGGATGACACAGCCAGCTATCCAGCCAATC 744
 QY 181 AAGAAGCACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACGCGATTGG 240
 Db 745 AAGAAGCACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACGCGATTGG 804
 QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGCGTGGCTGGCCCAACAACCTCTCGTGTGGCCAAA 300
 Db 805 CTGAAGGAGAAAGTGCAGAGCTTGCAGACACTGGCTGGCCCAACAACCTCTCGTGTGGCCAAA 864
 QY 301 GCCACAACGACACCTCGGAGGATATGAACGACCTCAACTCACTTCACAGCTCAGATG 360
 Db 865 GCCACAACGACACCTCGGAGGATATGAACGACCTCAACTCACTTCACAGCTCAGATG 924
 QY 361 GAGAACATCACCCTATCTCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCAGGACTTA 420
 Db 925 GAGAACATCACCCTATCTCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCAGGACTTA 984
 QY 421 CACAAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACAACCTGGAGGAAGCTTCCAG 480
 Db 985 CACAAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACAACCTGGAGGAAGCTTCCAG 1044
 QY 481 CTCTTTGACACGGATATTCTGACATCATTAGCAATATCAGTTACACAGCCCAACCTG 540
 Db 1045 CTCTTTGACACGGATATTCTGACATCATTAGCAATATCAGTTACACAGCCCAACCTG 1104
 QY 541 CGGACGCTCACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAAA 600
 Db 1105 CGGACGCTCACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAAA 1164
 QY 601 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCGGTTGGATTCGTGT 660
 Db 1165 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCGGTTGGATTCGTGT 1224
 QY 661 TCTCTCAGGATGCACAAAGATTTGATGAGGTGAGGCTAGACACTGAAGTAGCCAACTTA 720
 Db 1225 TCTCTCAGGATGCACAAAGATTTGATGAGGTGAGGCTAGACACTGAAGTAGCCAACTTA 1284

Qy 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTTCAGCTCATCAAGAAAT 780
 Db 1285 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTTCAGCTCATCAAGAAAT 1344
 Qy 781 TTTACAATACATACAGGTCCACCGGGCCCCAGGGGTCCAAGAGTGACAGAGGATCCAG 840
 Db 1345 TTTACAATACATACAGGTCCACCGGGCCCCAGGGGTCCAAGAGTGACAGAGGATCCAG 1404
 Qy 841 GGACCCCTGGCCCAACTGGCAACAAAGGACAGAAAGAGAGAGAGGGGAGCTGGACCA 900
 Db 1405 GGACCCCTGGCCCAACTGGCAACAAAGGACAGAAAGAGAGAGAGGGGAGCTGGACCA 1464
 Qy 901 CCTGGCCCTGGGGGTGAGAGAGGCCAAATGGACAGCTGGTCCCGGAGAGCGTGGC 960
 Db 1465 CCTGGCCCTGGGGGTGAGAGAGGCCAAATGGACAGCTGGTCCCGGAGAGCGTGGC 1524
 Qy 961 GGCAAGGATCTAAGGCTCCACGGGCCCCAAGGCTCCCGTGGTCCCGTGGGAGGCC 1020
 Db 1525 GGCAAGGATCTAAGGCTCCACGGGCCCCAAGGCTCCCGTGGTCCCGTGGGAGGCC 1584
 Qy 1021 GGCCCTCAGGGCCCGCTGCGGACCCAGGCCCCCGGCCCCCAGGCAAGAGGAGGACTC 1080
 Db 1585 GGCCCTCAGGGCCCGCTGCGGACCCAGGCCCCCGGCCCCCAGGCAAGAGGAGGACTC 1644
 Qy 1081 CCGGCTCTCAGGGCCCGCTGCGGACCCAGGCCCCCGGCCCCCAGGCAAGAGGAGGACTC 1140
 Db 1645 CCGGCTCTCAGGGCCCGCTGCGGACCCAGGCCCCCGGCCCCCAGGCAAGAGGAGGACTC 1704
 Qy 1141 GTGCTTGACCTCGGGGACTGCCAGGCTTGCTGGGTACACGAGCATCCAGGCCCAAG 1200
 Db 1705 GTGCTTGACCTCGGGGACTGCCAGGCTTGCTGGGTACACGAGCATCCAGGCCCAAG 1764
 Qy 1201 GGCCCCCGGCCCCCTCTGCGGACCCAGGAGGGGTGGTGGCCCTGCGGCTTCAGAAATGAG 1260
 Db 1765 GGCCCCCGGCCCCCTCTGCGGACCCAGGAGGGGTGGTGGCCCTGCGGCTTCAGAAATGAG 1824
 Qy 1261 CCAACCCCGGACCGGAGGACAAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAA 1320
 Db 1825 CCAACCCCGGACCGGAGGACAAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAA 1884
 Qy 1321 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAGAC 1380
 Db 1885 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAGAC 1944
 Qy 1381 AAGCTTTCACATCTTGTGTTTTCATTAACACTAGAGAGGAAACAGCAATGGATGATAAAAAACAG 1440
 Db 1945 AAGCTTTCACATCTTGTGTTTTCATTAACACTAGAGAGGAAACAGCAATGGATGATAAAAAACAG 2004
 Qy 1441 ATGGTAGGAGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAATGAATGG 1500
 Db 2005 ATGGTAGGAGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAATGAATGG 2064
 Qy 1501 AAGTGGCTGGATGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACCGCGATAC 1560
 Db 2065 AAGTGGCTGGATGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACCGCGATAC 2124
 Qy 1561 TGGGTCTATGCGCATGGCCAGGAGAACTGTGCTGGGTGATTTATGCTGGGCGAGTGG 1620
 Db 2125 TGGGTCTATGCGCATGGCCAGGAGAACTGTGCTGGGTGATTTATGCTGGGCGAGTGG 2184
 Qy 1621 AACGATTTCCAAATGTGAAGAGCTCAATAAATTTTCATTTTCGAAAAAGACAGGAGACAGTA 1680
 Db 2185 AACGATTTCCAAATGTGAAGAGCTCAATAAATTTTCATTTTCGAAAAAGACAGGAGACAGTA 2244
 Qy 1681 CTGTCATCTGCATTA 1695
 Db 2245 CTGTCATCTGCATTA 2259

RESULT 4
 AAC66903
 ID AAC66903 standard; cDNA; 2641 BP.

XX AAC66903;
 XX AC 27-MAR-2001 (first entry)
 XX DT Human EXMAD-14 coding sequence SEQ ID NO: 39.
 XX DE Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 XX KW inflammation; reproductive disorder; cardiovascular disorder;
 KW immune disorder; musculoskeletal disorder; developmental disorder;
 KW gastrointestinal disorder; cell proliferation disorder; ss.
 XX OS Homo sapiens.
 XX PN WO200068380-A2.
 XX PD 16-NOV-2000.
 XX 10-MAY-2000; 2000WO-US12811.
 PF 11-MAY-1999; 99US-0133643.
 PR 23-AUG-1999; 99US-0150409.
 XX (INCY-) INCYTE GENOMICS INC.
 PA Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 PI Azimzai Y;
 XX WPI: 2001-007395/01.
 DR P-FSDB; AAB27236.
 XX Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX Claim 4; Page 121-122; 129pp; English.
 PS The present invention provides the protein and coding sequences for 25
 XX novel extracellular matrix and adhesion-associated proteins (EXMADs).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 XX Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;
 SQ Query Match 99.9%; Score 1693.4; DB 22; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1694; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GTCAGCAATCTGCAGCAAGATACACGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 60
 Db 243 GTCAGCAATCTGCAGCAAGATACACGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 302
 Qy 61 TCTCATATATGGTTCATCATGNAACCTCAACACCTGAACCTGACCCAGGTGCAGCAGG 120
 Db 303 TCTCATATATGGTTCATCATGNAACCTCAACACCTGAACCTGACCCAGGTGCAGCAGG 362
 Qy 121 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCCAGGCAATC 180
 Db 363 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCCAGGCAATC 422
 Qy 181 AAGACGACTTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGACACGAGGATTGG 240
 Db 423 AAGACGACTTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGACACGAGGATTGG 482
 Qy 241 CTGAAGAGAAAGTGCAGAGCTTGCAGAGCTTGCAGAGCTGGCTGCCAACAACTCTGCGTTGCCCAA 300
 Db 483 CTGAAGAGAAAGTGCAGAGCTTGCAGAGCTTGCAGAGCTGGCTGCCAACAACTCTGCGTTGCCCAA 542

QY 1561 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATCTGGGCACTGG 1620
 Db 2157 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATCTGGGCACTGG 2216
 QY 1621 AAGGATTTCAATGTCAGACGCTCAATCTTTCATTTGCGAAAAGACAGGGAGACAGTA 1680
 Db 2217 AAGGATTTCAATGTCAGACGCTCAATCTTTCATTTGCGAAAAGACAGGGAGACAGTA 2276
 QY 1681 CTGTCATCTGCATTA 1695
 Db 2277 CTGTCATCTGCATTA 2291

RESULT 6

AAV55746

ID AAV55746 standard; cDNA; 2930 BP.

XX AAV55746;

XX 23-MAR-1999 (first entry)

XX Human secreted protein clone bv227_1 coding sequence.

XX Secreted protein; human; nutritional supplements; immune stimulant;

XX Immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;

XX activin/inhibin; chemokinesis; haemostasis; thrombolysis;

XX receptor/ligand activity; anti-inflammatory; tumour inhibitor;

XX cadherin/tumour invasion suppressor; ds.

XX Homo sapiens.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 67..693

FT /*tag= a

XX WO9855614-A2.

XX 10-DEC-1998.

XX 01-JUN-1998; 98WO-US11210.

XX 29-MAY-1998; 98US-0087255.

PR 04-JUN-1997; 97US-0868696.

PR 04-JUN-1997; 97US-0868697.

PR 04-JUN-1997; 97US-0868698.

PR 04-JUN-1997; 97US-0868699.

PR 04-JUN-1997; 97US-0868900.

PR 04-JUN-1997; 97US-0869191.

PR 04-JUN-1997; 97US-0869192.

PR 04-JUN-1997; 97US-0869193.

PR 04-JUN-1997; 97US-0869194.

XX (GENY) GENETICS INST INC.

XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;

PI McCoy JM, Racie LA, Spaulding V, Treacy M;

XX WPI: 1999-059912/05.

DR P-PSDB; AAW73628.

XX New polynucleotides encoding secreted human proteins - derived from

PT human foetal brain, adult brain, foetal kidney, adult ovary, adult

PT retina, adult placenta or adult uterus cDNA libraries

XX Claim 26; Page 87-88; 127pp; English.

XX This sequence encodes a human secreted protein of the invention.

CC This DNA sequence was isolated from a human adult brain cDNA

CC library, and was designated clone bv227_1. The DNAs and proteins

CC are predicted to have biological activities which would make them

CC suitable for treating, preventing or ameliorating medical conditions in

CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 XX
 SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;

Query Match 99.2%; Score 1681.4; DB 20; Length 2930;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1693; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GTCAGAAATCTGCAGCAAGATACCAGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 60
 Db 598 GTCAGAAATCTGCAGCAAGATACCAGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 657
 QY 61 TCTCATAATGTGTCTATCATGAACCTCAACACCTGACCCAGGTGCAGCAGAGG 120
 Db 658 TCTCATAATGTGTCTATCATGAACCTCAACACCTGACCCAGGTGCAGCAGAGG 716
 QY 121 AACCTCATCAGAAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGCTATCCAGCGAAATC 180
 Db 717 AACCTCATCAGAAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGCTATCCAGCGAAATC 776
 QY 181 AAGAAGGACCTTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCCAAGAGCAGGATTTG 240
 Db 777 AAGAAGGACCTTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCCAAGAGCAGGATTTG 836
 QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAAACAACTCTGGTTGGCCAAA 300
 Db 837 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAAACAACTCTGGTTGGCCAAA 896
 QY 301 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTTATTCACAGTTCAGATG 360
 Db 897 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTTATTCACAGTTCAGATG 956
 QY 361 GAGAACATCACCACCTATCTCTCAAGCCCAAGAGCAGCAACCTGAAAGACCTTGCAGGACTTA 420
 Db 957 GAGAACATCACCACCTATCTCTCAAGCCCAAGAGCAGCAACCTGAAAGACCTTGCAGGACTTA 1016
 QY 421 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 480
 Db 1017 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 1076
 QY 481 CTCTTTGAGACGGATATTTGTGAACATCATTTAGCAATATCAGTTACACGCCACCACTTG 540
 Db 1077 CTCTTTGAGACGGATATTTGTGAACATCATTTAGCAATATCAGTTACACGCCACCACTTG 1136
 QY 541 CGGACGCTGACCAGCAATCTAAATGAAGTTCAGGACCACTTGCACAGATACCCCTTACCAA 600
 Db 1137 CGGACGCTGACCAGCAATCTAAATGAAGTTCAGGACCACTTGCACAGATACCCCTTACCAA 1196
 QY 601 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCAAACATCCGTTTGGATTCGTGT 660
 Db 1197 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCAAACATCCGTTTGGATTCGTGT 1256
 QY 661 TCTCTCAGGATCAACAAGATTTGTAGGTGAGGTTAGACACTGAAGTAGCCCAACTTA 720
 Db 1257 TCTCTCAGGATCAACAAGATTTGTAGGTGAGGTTAGACACTGAAGTAGCCCAACTTA 1316
 QY 721 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCATCAAGAAT 780
 Db 1317 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCATCAAGAAT 1376
 QY 781 TTTTACAATACTACAGGTCCTCCAGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCCAG 840
 Db 1377 TTTTACAATACTACAGGTCCTCCAGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCCAG 1436

Db 192 TCGACGAGCTTTTCTTCAAGCCAAAGAGGACACGGATTGGCTGAAGGAGAAAGTGCAGA 251
 QY 260 GCTTGACAGCGTGGCTGCCAACAACCTCTGCGTTGGCCAAAGCCCAACAACAGCACACCTCGG 319
 Db 252 GCTTGACAGACCTGGCTGCCAACAACCTCTGCGTTGGCCAAAGCCCAACAACAGCACACCTCGG 311
 QY 320 AGGATATGAACAGCCAGCTCAACTCATCTCAGAGTTCAGATGAGAGCAATCACCACATATCT 379
 Db 312 AGGATATGAACAGCCAGCTCAACTCATCTCAGAGTTCAGATGAGAGCAATCACCACATATCT 371
 QY 380 CTCACGCCAACGAGAGACCTTGAAGACCTCGAGACCTTACACAAAGATGACAGCAATA 439
 Db 372 CTCACGCCAACGAGAGACCTTGAAGACCTCGAGACCTTACACAAAGATGACAGCAATA 431
 QY 440 GAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGAGGGATATTG 499
 Db 432 GAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGAGGGATATTG 491
 QY 500 TGAACATCATTTAGCAATATCAGTTACACAGCCACACCTCGGGAGCGTGCACGCAATC 559
 Db 492 TGAACATCATTTAGCAATATCAGTTACACAGCCACACCTCGGGAGCGTGCACGCAATC 551
 QY 560 TAAATGAAGTCAAGGACCTTGCACAGATACCTTACCACAAACACACAGATGATCTGACCT 619
 Db 552 TAAATGAAGTCAAGGACCTTGCACAGATACCTTACCACAAACACACAGATGATCTGACCT 611
 QY 620 CCTTGAATTAATACCTCGGCCAATCCGTTTGGATTCTCTCTCAGGATGCAACAAG 679
 Db 612 CCTTGAATTAATACCTCGGCCAATCCGTTTGGATTCTCTCTCAGGATGCAACAAG 671
 QY 680 ATTTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTTGGAAGAAA 739
 Db 672 ATTTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTTGGAAGAAA 731
 QY 740 TGAAGCTAGTACACTCCAAAGCTGTCAGCTCATCAAGAATTTTACAATACTACAGGTC 799
 Db 732 TGAAGCTAGTACACTCCAAAGCTGTCAGCTCATCAAGAATTTTACAATACTACAGGTC 791
 QY 800 CACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCACAGGACCCCTCGGCCCAACG 859
 Db 792 CACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCACAGGACCCCTCGGCCCAACG 851
 QY 860 GCAACAGGACGACAAAGAGAGAGAGGGGAGCCTGGACCACTTGCCCTCGCGGTGAGA 919
 Db 852 GCAACAGGACGACAAAGAGAGAGAGGGGAGCCTGGACCACTTGCCCTCGCGGTGAGA 911
 QY 920 GAGGCCCCAATTGACACAGCTGTCCCGCGGAGAGCGTGGCGCAAGAGATCTAAAGGCT 979
 Db 912 GAGGCCCCAATTGACACAGCTGTCCCGCGGAGAGCGTGGCGCAAGAGATCTAAAGGCT 971
 QY 980 CCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTG 1039
 Db 972 CCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTG 1031
 QY 1040 GGGACCCAGCCCCCGGGCCCAACAGGCAAGAGGACTCCCGGCGCCTCAGGGCCCCC 1099
 Db 1032 GGGACCCAGCCCCCGGGCCCAACAGGCAAGAGGACTCCCGGCGCCTCAGGGCCCCC 1091
 QY 1100 CTGGCTTCCAGGACCTTACAGGACACCGTTGGGAGCGCTGGGTGCTTGACCTCGGGGAC 1159
 Db 1092 CTGGCTTCCAGGACCTTACAGGACACCGTTGGGAGCGCTGGGTGCTTGACCTCGGGGAC 1151
 QY 1160 TGCCAGGCTTGCTGGGTTACAGGATGCCAGGCCCCAGGCCCCCGCCCTCCCTG 1219
 Db 1152 TGCCAGGCTTGCTGGGTTACAGGATGCCAGGCCCCAGGCCCCCGCCCTCCCTG 1211
 QY 1220 GGGCATCAGAGGGTGTGGCTGGCCCTGGCCCTGCAGATGAGCCACCCCGGACCGGGAGG 1279
 Db 1212 GGGCATCAGAGGGTGTGGCTGGCCCTGGCCCTGCAGATGAGCCACCCCGGACCGGGAGG 1271
 QY 1280 ACAATGGCTGCCCGCTCACTGGAAGACTTACAGACAAATGCTACTATTTTTCAGTTG 1339
 Db 1272 ACATAGCTGCCCGCTCACTGGAAGACTTACAGACAAATGCTACTATTTTTCAGTTG 1331

QY 1340 AGAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTGTT 1399
 Db 1332 AGAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTGTT 1391
 QY 1400 TCATAAACACTAGAGAGGAGCAAGCAATGATATAAAAAACAGATGGTAGGAGAGAGGCC 1459
 Db 1392 TCATAAACACTAGAGAGGAGCAAGCAATGATATAAAAAACAGATGGTAGGAGAGAGGCC 1451
 QY 1460 ACTGGATCGCCTCAGAGACTCAGAGCTGAGAGCTGAAATGAATGGCTGGATGGGACAT 1519
 Db 1452 ACTGGATCGCCTCAGAGACTCAGAGCTGAGAGCTGAAATGAATGGCTGGATGGGACAT 1511
 QY 1520 CTCAGACTTACAAAATTTGAAAGCTGGACGCGGATAACTGGGCTCATGCGCATGGGC 1579
 Db 1512 CTCAGACTTACAAAATTTGAAAGCTGGACGCGGATAACTGGGCTCATGCGCATGGGC 1571
 QY 1580 CAGGAGAGACTGTGCTGGTTGATTATGCTGGGAGTGGACGATTTCCCATGTGAAG 1639
 Db 1572 CAGGAGAGACTGTGCTGGTTGATTATGCTGGGAGTGGACGATTTCCCATGTGAAG 1631
 QY 1640 ACGTCAATAACTTCATTTCGAAAAAGACAGAGGAGACAGTACTGTCTATCTGCATTA 1695
 Db 1632 ACGTCAATAACTTCATTTCGAAAAAGACAGAGGAGACAGTACTGTCTATCTGCATTA 1687

RESULT 8
 AAI58842
 ID AAI58842 standard; cdna; 2318 BP.
 XX AC AAI58842;
 XX DT 22-OCT-2001 (first entry)
 XX DE Human polynucleotide SEQ ID NO 1045.
 DE KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW leukinaemia; ss.
 XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB; AAM39686.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 1045; 10078pp; English.

xx The invention relates to human nucleic acids (AAI57798-AAI61369) and
cc the encoded polypeptides (AAM38642-AAM4213) with nootropic,
cc immunosuppressant and cytostatic activity. The polynucleotides are useful
cc in gene therapy. A composition containing a polypeptide or polynucleotide
cc of the invention may be used to treat diseases of the peripheral nervous
cc system, such as peripheral nervous injuries, peripheral neuropathy and
cc localised neuropathies and central nervous system diseases, such as
cc Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
cc lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
cc utilisation of the activities such as: Immune system suppression,
cc Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
cc and thrombolytic activity, cancer diagnosis and therapy, drug screening,
cc assays for receptor activity, arthritis and inflammation, leukaemias and
cc C.N.S disorders.
cc Note: The sequence data for this patent did not form part of the printed
cc specification.
xx
SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other:
Query Match 98.0%; Score 1661.8; DB 22; Length 2318;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 20 ATACCAGCGTGTCCAGGGCAATCTGCAGAACCAATATGTTCTCATATGTTGGTCATCA 79
Db 12 ATACCAGCGTGTCCAGGGCAATCTGCAGAACCAATATGTTCTCATATGTTGGTCATCA 71
QY 80 TGAACCTCAACAACCTGACCTGACCCAGGTGCACAGAGGAACCTCATCAAGATCTGC 139
Db 72 TGAACCTCAACAACCTGACCTGACCCAGGTGCACAGAGGAACCTCATCAAGATCTGC 131
QY 140 AGCGGTCTGTGTGACACAAAGCCAGGTATCCAGCAATCAAGAACGACTTTCAAAATC 199
Db 132 AGCGGTCTGTGTGACACAAAGCCAGGTATCCAGCAATCAAGAACGACTTTCAAAATC 191
QY 200 TCACGAGGTTTTCTTCAAGCCAAAGAGGACACGATGGTGAAGGAGAAGTGCAGA 259
Db 192 TCACGAGGTTTTCTTCAAGCCAAAGAGGACACGATGGTGAAGGAGAAGTGCAGA 251
QY 260 GCTTGCAGACGCTGGTGCACAACTCTGCGTTGGCCAAAGCCAAACACGACACCTGG 319
Db 252 GCTTGCAGACACTGGTGCACAACTCTGCGTTGGCCAAAGCCAAACACGACACCTGG 311
QY 320 AGGATATGAACAGCCAGCTCAACTCATTCACAGGTGAGTGGAGAACATCACCACATCT 379
Db 312 AGGATATGAACAGCCAGCTCAACTCATTCACAGGTGAGTGGAGAACATCACCACATCT 371
QY 380 CTCAA-GCCAAAGCAGCAACCTGAAGACCTGACGAGGACTTACAAAGATGCAGAAAT 438
Db 372 CTCAAAGGCCAACGACGAGCAACCTGAAGACCTGACGAGGACTTACAAAGATGCAGAAAT 431
QY 439 AGAAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAGCTCTTTGAGACGGATATT 498
Db 432 AGAAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAGCTCTTTGAGACGGATATT 491
QY 499 GTGAACATCATTTAGCAATATCAGTTACACAGCCCAACCTCTGGACGCTGACCAAGCAAT 558
Db 492 GTGAACATCATTTAGCAATATCAGTTACACAGCCCAACCTCTGGACGCTGACCAAGCAAT 551
QY 559 CTAATGAAGTCAGACACACTTGCACAGATACCCCTTACCAAAACACACAGATGATCTGACC 618
Db 552 CTAATGAAGTCAGACACACTTGCACAGATACCCCTTACCAAAACACACAGATGATCTGACC 611
QY 619 TCCTTGAATAATACCTGGCCCAACCTCCGTTTGGATTCGTTCTCTCAGGATGCAACAA 678
Db 612 TCCTTGAATAATACCTGGCCCAACCTCCGTTTGGATTCGTTCTCTCAGGATGCAACAA 671
QY 679 GATTTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTTATGAAGAA 738
Db 672 GATTTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTTATGAAGAA 731
QY 739 ATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAAGT 798

Db 732 ATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACAATACTACAAGT 791
QY 799 CCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGCCCCAACT 858
Db 792 CCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGCCCCAACT 851
QY 859 GGCACAAGGGGACAGAAAGAGAGAGAGGGAGCCCTGGACCCTCGCCCTGCGGTGAG 918
Db 852 GGCACAAGGGGACAGAAAGAGAGAGAGGGAGCCCTGGACCCTGCGGTGAG 911
QY 919 AGAGGCCCAATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGATCTAAAGGC 978
Db 912 AGAGGCCCAATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGATCTAAAGGC 971
QY 979 TCCAGGGCCCCCAAGGGCTCCCGTGGTTCCTGGGAAGCCCGGCCCTCAGGGCCCCAGT 1038
Db 972 TCCAGGGCCCCCAAGGGCTCCCGTGGTTCCTGGGAAGCCCGGCCCTCAGGGCCCCAGT 1031
QY 1039 GGGGACCCAGGCCCCCGGGCCCAACAGAGCAAGAGAGGACTCCCGCGGCCCTCAGGGCCCCT 1098
Db 1032 GGGGACCCAGGCCCCCGGGCCCAACAGAGCAAGAGAGGACTCCCGCGGCCCTCAGGGCCCCT 1091
QY 1099 CTTGGCTTCCAGGAGCTTCAGGCGACCGTTGGGGAGCCCTGGGGTGCTGGACCTCGGGGA 1158
Db 1092 CTTGGCTTCCAGGAGCTTCAGGCGACCGTTGGGGAGCCCTGGGGTGCTGGACCTCGGGGA 1151
QY 1159 CTGCCAGGCTTCCTGGGGTACCGAGCATGCGCAGGCCCAAGGGCCCCCGGCCCTCCT 1218
Db 1152 CTGCCAGGCTTCCTGGGGTACCGAGCATGCGCAGGCCCAAGGGCCCCCGGCCCTCCT 1211
QY 1219 GGGCCATCAGGAGCGGTGGTGGCCCTGCGCCCTGCAAGTGAAGCAACCCCGGCGAGGAG 1278
Db 1212 GGGCCATCAGGAGCGGTGGTGGCCCTGCGCCCTGCAAGTGAAGCAACCCCGGCGAGGAG 1271
QY 1279 GACATGGCTGCCGCTCAGCTGGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTT 1338
Db 1272 GACATGGCTGCCGCTCAGCTGGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTT 1331
QY 1339 GAGAAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAAAGTCTTCAATCTTGT 1398
Db 1332 GAGAAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAAAGTCTTCAATCTTGT 1391
QY 1399 TTTATAAACACTAGAGAGGAACAGCAATGGATGATAAAAAACAGATGTAGGGAGAGAGC 1458
Db 1392 TTTATAAACACTAGAGAGGAACAGCAATGGATGATAAAAAACAGATGTAGGGAGAGAGC 1451
QY 1459 CACTGGATCGGCTCAGACTCAGAGCTGAGAGCGTGAAGTGAAGTGGCTGGATGGGACA 1518
Db 1452 CACTGGATCGGCTCAGACTCAGAGCTGAGAGCGTGAAGTGAAGTGGCTGGATGGGACA 1511
QY 1519 TCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTTGGGGTTCATGGCCATGG 1578
Db 1512 TCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTTGGGGTTCATGGCCATGG 1571
QY 1579 CCAGGAGAAAGACTGTGCTGGTTGATTATGCTGGCAGTGGGAACGATTTCCAATGTAA 1638
Db 1572 CCAGGAGAAAGACTGTGCTGGTTGATTATGCTGGCAGTGGGAACGATTTCCAATGTAA 1631
QY 1639 GACGTCAATACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCATCTGCATTA 1695
Db 1632 GACGTCAATACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCATCTGCATTA 1688

RESULT 9
ABL95574
ID ABL95574 standard; cDNA; 2005 BP.
xx
AC ABL95574;
xx
DT 19-JUL-2002 (first entry)
xx
DE Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.

XX	Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW	atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW	cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;
KW	antiarteriosclerotic; gene; ss.
XX	
OS	Homo sapiens.
XX	
XX	WO200208284-A2.
PN	
XX	
PD	31-JAN-2002.
XX	
PF	09-JUL-2001; 2001WO-US21735.
XX	
XX	20-JUL-2000; 2000US-219556P.
PR	25-JUL-2000; 2000US-220624P.
PR	25-JUL-2000; 2000US-220664P.
PR	28-JUL-2000; 2000WO-US20710.
PR	02-AUG-2000; 2000US-222695P.
PR	17-AUG-2000; 2000US-0643657.
PR	23-AUG-2000; 2000WO-US23522.
PR	24-AUG-2000; 2000WO-US23328.
PR	07-SEP-2000; 2000US-230378P.
PR	15-SEP-2000; 2000US-000000P.
PR	18-SEP-2000; 2000US-0664610.
PR	18-SEP-2000; 2000US-0665350.
PR	24-OCT-2000; 2000US-242922P.
PR	08-NOV-2000; 2000US-0709238.
PR	08-NOV-2000; 2000WO-US30952.
PR	10-NOV-2000; 2000WO-US30873.
PR	01-DEC-2000; 2000WO-US32678.
PR	20-DEC-2000; 2000US-0747259.
PR	20-DEC-2000; 2000WO-US34956.
PR	22-JAN-2001; 2001US-0767609.
PR	28-FEB-2001; 2001US-0796498.
PR	28-FEB-2001; 2001WO-US06520.
PR	01-MAR-2001; 2001WO-US06666.
PR	09-MAR-2001; 2001US-0802706.
PR	14-MAR-2001; 2001US-0808689.
PR	22-MAR-2001; 2001US-0816744.
PR	05-APR-2001; 2001US-0828366.
PR	10-MAY-2001; 2001US-0854208.
PR	10-MAY-2001; 2001US-0854280.
PR	25-MAY-2001; 2001US-0866028.
PR	25-MAY-2001; 2001US-0866034.
PR	25-MAY-2001; 2001WO-US17092.
PR	30-MAY-2001; 2001US-0870574.
PR	30-MAY-2001; 2001WO-US17443.
PR	01-JUN-2001; 2001WO-US17800.
PR	20-JUN-2001; 2001WO-US19692.
PR	28-JUN-2001; 2001WO-US00000.
XX	
PA	(GETH) GENENTECH INC.
PA	{BAKE/} BAKER K P.
PA	{FERR/} FERRARA N.
PA	{GERB/} GERBER H.
PA	{GERR/} GERRITSEN M E.
PA	{GODD/} GODDARD A.
PA	{GODO/} GODOWSKI P J.
PA	{GURN/} GURNEY A L.
PA	{HILL/} HILLAN K J.
PA	{MARS/} MARSTERS S A.
PA	{PANJ/} PAN J.
PA	{PAON/} PAONI N F.
PA	{STEP/} STEPHAN J F.
PA	{WATA/} WATANABE C K.
PA	{WILL/} WILLIAMS P M.
PA	{WOOD/} WOOD W I.
XX	
PI	Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI	Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI	Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX	

Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

RESULT 10	
ABL88085	
ID	ABL88085 standard; cDNA; 2005 BP.
XX	
XX	ABL88085;
XX	
XX	16-MAY-2002 (first entry)
DT	
XX	
XX	
DE	Human PRO7223 cDNA sequence SEQ ID NO:27.
XX	
KW	Human; angiogenesis; cardiac; cytostatic; antiangiogenic; hypotensive;
KW	vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW	gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW	angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW	age-related macular degeneration; arterial stenosis; angina;
KW	rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW	lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW	wound healing; chromosome mapping; gene mapping; gene; ss.
XX	
XX	Homo sapiens.
OS	
XX	

Query Match 87.2%; Score 1477.6; DB 24; Length 2005;									
Best Local Similarity 99.7%; Pred. No. 0;									
Matches 1491; Conservative 0; Mismatches 4; Indels 1; Gaps 1;									
QY	200	TGCAGCAGGTTTCTTCAAGCCAAAGACACGAGTGGTGAAGAGAAAGTGCAGA	259						
Db	9	TCGGCAGGTTTCTTCAAGCCAAAGACACGAGTGGTGAAGAGAAAGTGCAGA	68						
QY	260	GCTTGACAGCGCTGGCTGCCAACACTCTGCGTTGGCCAAAGCCCAACAGCACCCCTCG	319						
Db	69	GCTTGACAGACACTGGCTGCCAACACTCTGCGTTGGCCAAAGCCCAACAGCACCCCTCG	128						
QY	320	AGGATATGAACGCCAGCTCAACTCATTCACAGGTCAGATGGAGAACATCACCACATCT	379						
Db	129	AGGATATGAACGCCAGCTCAACTCATTCACAGGTCAGATGGAGAACATCACCACATCT	188						
QY	380	CTCAAGCCAAAGAGAGAACCTTGAAGACCTGAGAGCTTACACAAAGATGCAGAAATA	439						
Db	189	CTCAAGCCAAAGAGAGAACCTTGAAGACCTGAGAGCTTACACAAAGATGCAGAAATA	248						
QY	440	GAACAGCCATCAAGTTCAACCACTGAGGAGACGCTTCCAGCTCTTTGAGAGCGATATTG	499						
Db	249	GAACAGCCATCAAGTTCAACCACTGAGGAGACGCTTCCAGCTCTTTGAGAGCGATATTG	308						
QY	500	TGAACATCATATTAGCAATATCAGTTACACAGCCACACCTGCGGACGCTGACCAATC	559						
Db	309	TGAACATCATATTAGCAATATCAGTTACACAGCCACACCTGCGGACGCTGACCAATC	368						
QY	560	TAAATCAAGTCAGGACCACTTGCACAGATACCTTACCAAAACACACAGATGATCTACCT	619						
Db	369	TAAATCAAGTCAGGACCACTTGCACAGATACCTTACCAAAACACACAGATGATCTACCT	427						
QY	620	CCTTGAATAATACCTGGCCAAACATCCGTTTGGATTCTGTCTCTCAGGATGCAACAAG	679						
Db	428	CCTTGAATAATACCTGGCCAAACATCCGTTTGGATTCTGTCTCTCAGGATGCAACAAG	487						
QY	680	ATTGTATGAGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAA	739						
Db	488	ATTGTATGAGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAA	547						
QY	740	TGAAGCTAGTACACTCAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAGGTC	799						
Db	548	TGAAGCTAGTACACTCAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAGGTC	607						
QY	800	CACCGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGACCCCTGGGCCCAACTG	859						
Db	608	CACCGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGACCCCTGGGCCCAACTG	667						
QY	860	GCNACAGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACTTGGCCCTGGCGGTGAGA	919						
Db	668	GCNACAGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACTTGGCCCTGGCGGTGAGA	727						
QY	920	GAGGCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGCAAGAGATCTAAAGGCT	979						
Db	728	GAGGCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGCAAGAGATCTAAAGGCT	787						
QY	980	CCAGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGGGCCCCAGTG	1039						
Db	788	CCAGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGGGCCCCAGTG	847						
QY	1040	GGGACCCAGGCCCCCGGGCCCAACAGGCAAAAGAGGACTCCCGGCCCTCAGGGCCCTC	1099						
Db	848	GGGACCCAGGCCCCCGGGCCCAACAGGCAAAAGAGGACTCCCGGCCCTCAGGGCCCTC	907						
QY	1100	CTGGCTTCAGGACACTTCAAGGCACCGTTGGGAGCGCTGGGTCGTGGACCTCGGGAC	1159						
Db	908	CTGGCTTCAGGACACTTCAAGGCACCGTTGGGAGCGCTGGGTCGTGGACCTCGGGAC	967						
QY	1160	TGCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCCCTG	1219						
Db	968	TGCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCCCTG	1027						

QY	1220	GCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCAAACCCGGCACCGGAGG	1279						
Db	1028	GCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCAAACCCGGCACCGGAGG	1087						
QY	1280	ACAATGGCTGCCCGCTCACTGGAAGAATTCACACACAAATGCTACTATTTTCAGTTG	1339						
Db	1088	ACAATAGCTGCCCGCTCACTGGAAGAATTCACACACAAATGCTACTATTTTCAGTTG	1147						
QY	1340	AGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTT	1399						
Db	1148	AGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTT	1207						
QY	1400	TCATAAACACTAGAGAGGAACAGCAATGGATGATAAAAAACAGATGGTAGGAGAGAGAGCC	1459						
Db	1208	TCATAAACACTAGAGAGGAACAGCAATGGATGATAAAAAACAGATGGTAGGAGAGAGAGCC	1267						
QY	1460	ACTGGATCGGCTCAGAGACTCAGAGCGTGAATGAATGGAAGTGGTGGATGGGACAT	1519						
Db	1268	ACTGGATCGGCTCAGAGACTCAGAGCGTGAATGAATGGAAGTGGTGGATGGGACAT	1327						
QY	1520	CTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGATAACTGGGGTCTATGGCCATGGGC	1579						
Db	1328	CTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGATAACTGGGGTCTATGGCCATGGGC	1387						
QY	1580	CAGGAGAAAGACTGTGCTGGTTGATTTATGCTGGGCGAGTGGAAACGATTTCCAATGTGAAG	1639						
Db	1388	CAGGAGAAAGACTGTGCTGGTTGATTTATGCTGGGCGAGTGGAAACGATTTCCAATGTGAAG	1447						
QY	1640	AGCTCAATAACTTCATTTCCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA	1695						
Db	1448	AGCTCAATAACTTCATTTCCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA	1503						

RESULT 11

AAH43037

ID AAH43037 standard; cDNA; 2637 BP.

XX AC AAH43037;

XX AC AAH43037;

DE Nucleotide sequence of a human scavenger receptor.

KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity; arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;

KW low density lipoprotein; ss.

OS Homo sapiens.

Key Location/Qualifiers

FT CDS 92..2320

FT /*tag= a

FT /product= "scavenger receptor"

XX WO200159107-A1.

PN 16-AUG-2001.

XX 08-FEB-2001; 2001WO-JP00874.

XX 14-FEB-2000; 2000JP-0035155.

PR 10-OCT-2000; 2000JP-0309068.

XX (FUSO) FUSO PHARM IND LTD.

XX Wakamiya N;

XX WPI; 2001-497076/54.

DR P-PSDB; AAG63347.

XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure, useful for treatment and diagnosis of diseases associated with oxidized low-density lipoprotein accumulation

XX PS Claim 6; Page 88-93; 118pp; Japanese.

XX CC The present sequence encodes a human scavenger receptor, designated

CC SRC1-p1. The SRC1-p1 polypeptide has a collectin-like structure. They

CC are useful in clarifying the functions of macrophages and basal

CC immunity. They are also useful in the treatment, prevention, diagnosis

CC and investigation of diseases such as arteriosclerosis, diabetic

CC complications, bacterial infection and restenosis following angioplasty,

CC which are associated with accumulation of oxidized low density

CC lipoprotein and the binding of advanced glycation end-products into

CC cells.

XX SQ Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;

Query Match 80.2%; Score 1359; DB 22; Length 2637;

Best Local Similarity 87.6%; Pred. No. 0;

Matches 1485; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 GTCAGCAATCTGCAGCAAGATACACGCGTGTCCAGGGCAATCTGCAGAACCAATGTAT 60

DB 623 GTCAAAATCTGCAACAGATACTAGTGTCTCCAGGGCAATCTGCAGAGCCAAATGTAT 682

QY 61 TCTCATATGTGGTTCATCATGAACCTCAACAACCTGAACCTACCCAGGTGCAGCAGAGG 120

DB 683 TCTCAGAGCGTGGTTCATGAACCTCAACAACCTGAACCTTAACCCAGGTTCAGCAGAGG 742

QY 121 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACAAGCCAGGCTATCCAGCGAATC 180

DB 743 AACCTTATCTCAATCTGCAGCAGTCTGTGGATGACACAAGCCGTGCCATCCAGCAAT 802

QY 181 AAGAACGACTTTCAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGACACGATTGG 240

DB 803 AAGAATGATTTCCAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGACACGATTGG 862

QY 241 CTGAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACCTCTGCGTTGGCCAAA 300

DB 863 CTAAAGGAAAAAGTACAGAGCTTGACAGCATTTGGTGCACAACCTCTGCCCTGGCCAAA 922

QY 301 GCCAACAACGACACCTTGAGGATATGAACAGCGAGCTCAACTCATCTACAGGTCAGATG 360

DB 923 GCCAACAATGACACCTTAGAGGATATGAATAGCCAGCTCAGCTCATCTACAGGTCAGATG 982

QY 361 GAGAACATACACATCTCTCAAGCCCAACGAGCAGAACCTTGAAGACCTGCGAGCACTTA 420

DB 983 GACAACTTACCACTATCTCAAGCCCAACGAGCAGAGGCTTGAAGACCTTTCAGGACTTA 1042

QY 421 CACAAGATGCAGAAATAGACAGCGATCAAGTTCAACCAACTGGAGAACCTTCCAG 480

DB 1043 CACAAGGATACAGAAATAGACAGCTGTCAAGTTTCAGCCAACTTGAGAACGCTTCCAG 1102

QY 481 CTCCTTGAGACGGATATGTGAACATCATTTAGCAATATCAGTTACACAGCCCAACCTG 540

DB 1103 GTCTTTGAGACAGATATGTGAACATCATTTAGCAATATCAGTTACACAGCCCATCACCTG 1162

QY 541 CGGACCTGACAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA 600

DB 1163 AGGACACTGACCAGCAATCTGAATGATTTAGGACCACTGCACAGACCTTGACCAGA 1222

QY 601 CACACAGATGATCTGACCTCTTGAATTAATACCTCGGCCAACAATCGGTTTGATCTGTT 660

DB 1223 CACAGGGATGACCTGACCTCTTGAATTAATACCACTAGTCAACATCGGCTTGGATCTATT 1282

QY 661 TCTCTCAGGATCAACAAGATTTGATGAGTTCGAGGTAGACACTCAAGTACCCCACTTA 720

DB 1283 TCTCTCAGGATGCAGACAGATGATGAGGTCAAAAGTTAGACACTGAAGTGGCCCACTTA 1342

QY 721 TCAGTGATTTGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAT 780

DB 1343 TCAGTGGTTTGAAGAGATGAATGTTGACTCCAAAGCAGGTCAGCTCATCAAGAAC 1402

QY 781 TTATCAATATCTACAGGTCCACGGGCCCGAGGTTCCAAAGAGGTGACAGAGATCCAG 840

Db 1403 TTTACCATTTCTACAAGGTCTCTCTGGCCCGCAGAGGTCTCAAAAAGGTGACAGAGGATCTCAG 1462

QY 841 GGACCCCTTGGCCCAACTGGCAACAGAGGACAGAAAGAGAGAGAGGAGGAGCCTGACCA 900

Db 1463 GGACCACTTGGTCCCACTGGCAACAAGGACAGAAAGAGAGAGAGGAGCCTGGTCCA 1522

QY 901 CTGGCCCTTGGGGTGAGAGAGGCCCAATTTGACACAGCTGGTCCCGCCGGAGAGCGTGGC 960

Db 1523 CCTGGCCCTTGGGGTGAGAGAGGCCCAATTTGACACAGCTGGTCCCGCCCTCTGGAGAGCGTGGC 1582

QY 961 GCAAGAGGATCTAAAGGCTCCAGGGCCCAAGGCTCCCGTGGTTCCTTCCCTGGAGAGCC 1020

Db 1583 AGCAAGGATCCAAAGGCTCACAGGCTCCCAAGGATCTCGTGGTCCCGCAGGAGAGCCT 1642

QY 1021 GGCCCTCAGGGCCCGCAGTGGGACCCAGGCCCCCGGGCCCGCAGGCAAGAGGAGCCTC 1080

Db 1643 GGCCCTCAGGACCTAGTGGGACCCAGGACCCAGGCTCCAGGCAAGGATGAGCCTC 1702

QY 1081 CCGGCGCTCAGGGCCCTCTGGTTCAGGACCTTTCAGGACCTTTCAGGACCTTTCAGGAGCCTGGG 1140

Db 1703 CCTGGCCCTCAGGGCCCTCTGGTTCAGGACCTTTCAGGACCTTTCAGGAGCCTGGG 1762

QY 1141 GTGCTTGGACCTCGGGGCTGCGGCTTGCCTGGGTACCGGCTCCAGGCCCCCAAG 1200

Db 1763 GTACCTGGACCTCGGGGCTTGCAGGCTTGCAGGGGTGCGGAGGATGCTTGGGCTAAG 1822

QY 1201 GGCCCGCCCGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAG 1260

Db 1823 GGACCACTTGGCCCTCCAGGCCCCCTCAGGACCAATGGAGCCATTTGGCTCTGCAGAAATGA 1882

QY 1261 CCAACCCCGCCCGGAGGACCAATGGCTGCCCGCCCTCAGTGAAGAACTTTCACAGACAA 1320

Db 1883 CCAACCCCGGAGATCAGAGGTCAACGGATCTCCGCTCAGTGAAGAACTTTCACAGATAA 1942

QY 1321 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTGTGAAGAC 1380

Db 1943 TGCTACTATTTTTCATTTGAAAAAGAAATTTTGAAGATGCTAAGCTTTCTGTGAAGAC 2002

QY 1381 AAGTCTTCACATCTTGTCTTTCATTAACACTAGAGAGAACAGCAATGGATAAAAAACAG 1440

Db 2003 AATCTTCCATCTCTGTTTTCATTAACACTAGAGAGAACAGCAATGGATAAAAAGCAT 2062

QY 1441 ATGGTAGGAGAGAGAGCCACTGGATCGCCCTCAGAGCTCAGAGCGTGAATGAATGG 1500

Db 2063 ACCGTGGGAGAGAGAGCCATTTGGATCGCCCTCAGAGCTCAGAGCAAGGAGCAATGG 2122

QY 1501 AAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCGCGATAC 1560

Db 2123 AAGTGGCTAGACGGGTCACTTGTATTACAAAACTGGAAAGCTGGCAACCAAGATAAC 2182

QY 1561 TGGGGTCAATGGCCATGGGCGCAGGAGAGACTGTGTGGTGTGATTTATCTGGGAGTGG 1620

Db 2183 TGGGCGAGTGGCCATGGGCGCAGGAGAGAGACTGTGTGGTGTGATTTACGAGGACAGTGG 2242

QY 1621 AACGATTTCCAATGTGAAGACGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACAGTA 1680

Db 2243 AATGACTTCCAGTGTGATGAATCAATAACTTTCATTTGTGAGAGAGAAAGGAGGAGCA 2302

QY 1681 CTGTCTATCTGCATTA 1695

Db 2303 CCATCATCATCATTA 2317

RESULT 12

AAH43054

ID AAH43054 standard; DNA; 2256 BP.

XX

XX AAH43054;

XX AC

XX 15-OCT-2001 (first entry)

XX

DE Nucleotide sequence of a human scavenger receptor.

XX

KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.

OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 74..1936
 FT /*tag= a
 FT /product= "scavenger receptor"

XX W0200159107-A1.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-JP00874.

XX 14-FEB-2000; 2000JP-0035155.

XX 10-OCT-2000; 2000JP-0309068.

XX (FUSO) FUSO PHARM IND LTD.

XX Wakamiya N;

XX WPI; 2001-497076/54.

XX P-PSDB; AAG63350.

XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 PT useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation

XX Claim 4; Page 105-109; 118pp; Japanese.

XX The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.

XX Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Query Match 55.5%; Score 941; DB 22; Length 2256;
 Best Local Similarity 78.1%; Pred. No. 5.1e-229;
 Matches 1323; Conservative 0; Mismatches 0; Indels 372; Gaps 1;

QY 1 GTCACGAATCTGCAGCAAGATACCAAGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 60
 DB |||||

DB 605 GTCACGAATCTGCAGCAAGATACCAAGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 654
 DB |||||

QY 61 TCTCATAATCTGTCATCATGACCTCACACCTCAACCTGACCCAGCTGCAGCAGAGG 120
 DB |||||

DB 665 TCTCATAATCTGTCATCATGACCTCACACCTCAACCTGACCCAGCTGCAGCAGAGG 724
 DB |||||

QY 121 AACCTCATACGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGCAATC 180
 DB |||||

DB 725 AACCTCATACGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGCAATC 784
 DB |||||

QY 181 AGAAGCACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGGACACGGATTGG 240
 DB |||||

DB 785 AAGAAGCACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGGACACGGATTGG 844
 DB |||||

QY 241 CTGAAGGAGAAAGTCAGAGCTGCAGACGCTGGCTGCGCAACAACTCTCGTTGGCCAAA 300
 DB |||||

DB 845 CTGAAGGAGAAAGTCAGAGCTGCAGACGCTGGCTGCGCAACAACTCTCGTTGGCCAAA 904
 DB |||||

QY 301 GCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAAGTG 360
 DB |||||

DB 905 GCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAAGTG 964
 DB |||||

QY 361 GAGAACATCACCACCTATCTCTCAAGCCAAAGCAGCAAGCACTGAAAGACCTGCAGGACTTA 420
 DB |||||

DB 965 GAGAACATCACCACCTATCTCTCAAGCCAAAGCAGCAAGCACTGAAAGACCTGCAGGACTTA 1024
 DB |||||

QY 421 CACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGACCTTCCAG 480
 DB |||||

DB 1025 CACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGACCTTCCAG 1084
 DB |||||

QY 481 CTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTTACACAGCCACCACTTG 540
 DB |||||

DB 1085 CTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTTACACAGCCACCACTTG 1144
 DB |||||

QY 541 CGGACGCTGACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 600
 DB |||||

DB 1145 CGGACGCTGACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 1204
 DB |||||

QY 601 CACACAGATGATCTGACCTCCTTGAATATACCTTGGCCAAACATCCGTTTGGATTCTGTT 660
 DB |||||

DB 1205 CACACAGATGATCTGACCTCCTTGAATATACCTTGGCCAAACATCCGTTTGGATTCTGTT 1264
 DB |||||

QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTCAAGTAGGCAACTTA 720
 DB |||||

DB 1265 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTCAAGTAGGCAACTTA 1324
 DB |||||

QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCCAAGCATGGTCAGCTCATCAAGAAT 780
 DB |||||

DB 1325 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCCAAGCATGGTCAGCTCATCAAGAAT 1384
 DB |||||

QY 781 TTTACAATACTACAAGGTCCACCGGCCCCAGGGGTCCTCAAGAGGTACACAGATCCCCAG 840
 DB |||||

DB 1385 TTTACAATACTACAAGGTCCACCGGCCCCAGGGGTCCTCAAGAGGTACACAGATCCCCAG 1444
 DB |||||

QY 841 GGACCCCTCGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCA 900
 DB |||||

DB 1445 GGACCCCTCGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCA 1504
 DB |||||

QY 901 CTTGGCCCTCGGGGTGAGAGAGGCCCAATTTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 960
 DB |||||

DB 1505 CTTGGCCCTCGG----- 1516
 DB |||||

QY 961 GGCAAAGGATCTAAAGGCTCCAGGGCCCAAGGCTCCCGTGGTTCCCTCGGGAAGCCC 1020
 DB |||||

DB 1517 ----- 1516
 DB |||||

QY 1021 GGCCCTCAGGGCCCTCAGTGGGACCCAGGCCCCCGGCCCCAGGCAAGAGGACTC 1080
 DB |||||

DB 1517 ----- 1516
 DB |||||

QY 1081 CCGGGCCCTCAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGACCGTTGGGGAGCGCTGGG 1140
 DB |||||

DB 1517 ----- 1516
 DB |||||

QY 1141 GTGCCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAG 1200
 DB |||||

DB 1517 ----- 1516
 DB |||||

QY 1201 GGCCCCCGGCCCTCCTGGCCCATCAGGAGGGTGGTGGCCCTGGCCCTGCAGAGATGAG 1260
 DB |||||

DB 1517 ----- 1516
 DB |||||

QY 1261 CCAACCCCGGACCGGAGGACAAATGGCTGCCGCCCTCCTGGAAGAACTTCACAGACAAA 1320
 DB |||||

DB 1517 ----- 1552
 DB |||||

QY 1321 TGCTACTATTTTTCAGTTTCAGAAAGAAATTTTGGAGATGCAAGGCTTTTCTGTGAAGAC 1380
 DB |||||

DB 1553 TGCTACTATTTTTCAGTTTCAGAAAGAAATTTTGGAGATGCAAGGCTTTTCTGTGAAGAC 1612
 DB |||||

QY 1381 AAGTCTTTCACATCTTGTTTTTCATTAACACTAGAGAGGAACAGCAATGGATAAAAAACAG 1440
 DB |||||

DB 1613 AAGTCTTTCACATCTTGTTTTTCATTAACACTAGAGAGGAACAGCAATGGATAAAAAACAG 1672
 DB |||||

QY 1441 ATGGTAGGGAGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGAGCGGTGAAAAATGAATGG 1500
 DB |||||

|||||
Db 1673 ATGGTAGGAGAGAGCCACTGGATCGCCTCAGACACTCAGACGCTGAAATGAATGG 1732
Qy 1501 AAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCGGATAC 1560
Db 1733 AAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCGGATAC 1792
Qy 1561 TGGGCTCATGCGCATATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGGAGTGG 1620
Db 1793 TGGGCTCATGCGCATATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGGAGTGG 1852
Qy 1621 AACGATTTCCATGTGAAGACGCTCAATTAACCTTTCGGAAGAGAGAGGAGACAGTA 1680
Db 1853 AACGATTTCCATGTGAAGACGCTCAATTAACCTTTCGGAAGAGAGAGGAGACAGTA 1912
Qy 1681 CTGTCATCTGCATTA 1695
Db 1913 CTGTCATCTGCATTA 1927
RESULT 13
AAS71133
ID AAS71133 standard; cDNA: 1521 BP.
XX AC AAS71133;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #6937.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG06946.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 1; SEQ ID No 6937; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1521 BP; 386 A; 420 C; 425 G; 290 T; 0 other;
Query Match 52.1%; Score 883.8; DB 23; Length 1521;
Best Local Similarity 99.8%; Pred. No. 1.5e-214;
Matches 885; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 793 CAAGGTCCACCGGGCCCGCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGCCCTCGGC 852
Db 172 CAAGGTCCACCGGGCCCGCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGCCCTCGGC 231
Qy 853 CCACTGGCAACAGGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCCTGGCCTGGC 912
Db 232 CCACTGGCAACAGGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCCTGGCCTGGC 291
Qy 913 GGTGAGAGAGCCCAATTGGACCACTGTCCCGCCGAGAGCGTGGCGGCAAGAGATCT 972
Db 292 GGTGAGAGAGCCCAATTGGACCACTGTCCCGCCGAGAGCGTGGCGGCAAGAGATCT 351
Qy 973 AAAGCTCCCGAGGGCCCAAGGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGGGC 1032
Db 352 AAAGCTCCCGAGGGCCCAAGGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGGGC 411
Qy 1033 CCCAGTGGGGAGCCAGGCCCCCGGGCCACCAGAGCAAGAGAGGACTCCCGCGGCCCTCAG 1092
Db 412 TCCAGTGGGGAGCCAGGCCCCCGGGCCACCAGAGCAAGAGAGGACTCCCGCGGCCCTCAG 471
Qy 1093 GGCCTCTCTGCTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGGCTGGACCT 1154
Db 472 GGCCTCTCTGCTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGGCTGGACCT 531
Qy 1153 CGGGGACTGCCAGGCTTGCTGGGTACAGGCATGCCAGAGCCCAAGAGGCCCGCCCGGC 1212
Db 532 CGGGGACTGCCAGGCTTGCTGGGTACAGGCATGCCAGAGCCCAAGAGGCCCGCCCGGC 591
Qy 1213 CTTCTGGCCCATCAGAGAGCGGTGTGCCCTGGCCCTGCGAGAAAGAGCAACCCGGCA 1272
Db 592 CTTCTGGCCCATCAGAGAGCGGTGTGCCCTGGCCCTGCGAGAAAGAGCAACCCGGCA 651
Qy 1273 CGGAGAGCAATGCTGCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTATTTT 1332
Db 652 CGGAGAGCAATGCTGCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTATTTT 711
Qy 1333 TCAGTTGAGAAAGAAATTTTGAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCACAT 1392
Db 712 TCAGTTGAGAAAGAAATTTTGAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCACAT 771
Qy 1393 CTTGTTTTCTAATAACACTAGAGAGGAAACAGCAATGATGATAAAACAGATGTGAGGAGA 1452
Db 772 CTTGTTTTCTAATAACACTAGAGAGGAAACAGCAATGATGATAAAACAGATGTGAGGAGA 831
Qy 1453 GAGAGCCACTGGATCGGCTCAGACTCAGAGCTGGAATGAATGGAAGTGGCTGGAT 1512
Db 832 GAGAGCCACTGGATCGGCTCAGACTCAGAGCTGGAATGAATGGAAGTGGCTGGAT 891
Qy 1513 GGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCGGATGAATCTGGGTCATGGC 1572
Db 892 GGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCGGATGAATCTGGGTCATGGC 951
Qy 1573 CATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGCAGTGGAAACGATTTCCAA 1632
Db 952 CATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGCAGTGGAAACGATTTCCAA 1011
Qy 1633 TGTGAAGAGCTCAATAACTTTCATTTGCGAAAAGACAGGGAGACAGT 1679
Db 1012 TGTGAAGAGCTCAATAACTTTCATTTGCGAAAAGACAGGGAGACAGT 1058

[illegible]

XX Homo sapiens.
 OS W0909159-A1.
 PN XX
 PD 25-FEB-1999.
 XX
 PF 12-AUG-1998; 98WO-JP03602.
 XX
 PR 30-JUL-1998; 98JP-0230121.
 PR 13-AUG-1997; 97JP-0233396.
 XX
 PA (NIBS) JAPAN TOBACCO INC.
 XX
 PI Nakamura Y, Tokino T;
 XX WPI; 1999-181032/15.
 DR P-PSDB; RAY00992.
 DR
 XX Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress
 XX
 PS Claim 4; Page 119-127; 175pp; Japanese.
 XX
 CC This sequence encodes the human cellular stress response 1 (CSR1) protein
 CC of the invention. The CSR proteins are macrophage scavenger receptor
 CC proteins. The CSR proteins can be used in the treatment, gene therapy
 CC and diagnosis of diseases in which intracellular stress is important,
 CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
 CC infection. Expression of the proteins is induced in vivo in response to
 CC intracellular stress, and inhibits cell death as a result of such stress.
 XX
 SQ Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;
 Query Match 14.6%; Score 248.2; DB 20; Length 3810;
 Best Local Similarity 50.5%; Pred. No. 1e-52;
 Matches 601; Conservative 0; Mismatches 588; Indels 0; Gaps 0;
 QY 42 TCTGAGAACCAATGTTCTCATATGTTGTCATCATGTAACCTCAACAACTGAACT 101
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 890 TCTGAAGGACCTCAACCAAGAGTGGAGCGGTCAAGATGTCAGGCTGCAGTGCACCATCT 949
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 102 GACCCAGTGCAGAGAGAACCTCATCAGATCTGCAGCGGTCTGTGGATGACACAAAG 161
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 950 CACCGTGGGGCAGACTCCGAGTGGATCCACGGGATCCAGCGGAAGACAGAGGAGAC 1009
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 162 CCAGCTTCCAGCGAATCAAGAACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGC 221
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1010 CCGTACCCTCAGAGATGTCACCGACTGGCAGAACTACACACGGCTCTTCAGCGGCT 1069
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 222 CAAGAAGGACACGGATTTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAA 281
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1070 GCGCACCACTCCACCAAGACTGGAGCGGTCAAGAACATCCAGGCCACCTGGGGC 1129
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 282 CAACTCTGGTTGGCCAAAGCAACAGCACACCTCGAGGATATGACAGCCACTCAA 341
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1130 CTCCTCAACAGGCGATCAGCCAGAACTCAGAGAGATGCACGACCTGGTACTCCAGGTAT 1189
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 342 CTCATTACAGTGCAGATGAGAAATCACCCTATCTCTCAAGCAACAGCAGCAACCT 401
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1190 GGGCTTGCAGCTGCAGCTGGGATAATCTCGTCTCTCTGGATGACACAGAGAAAT 1249
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 402 GAAAGACCTGCAGGACTTACAAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCA 461
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1250 GCATGATCTTCAGTACCATACCCACTACGCCAGAACCGCACTGGGAGAGGTTTGAGTC 1309
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 462 ACTGGAGGAGCGTTCCAGCTCTTTGAGACGAGATATGTGAACATCATAGCAATATCAG 521
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1310 TCTGGAAGGAGCGATGGCTTCTCAGGAGATGAAATTTGGCACCATCTTCAACCAATCAA 1369
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 522 TTACACAGCCCAACCACTGGCGACGCTGACCAAGCAATCTAATGAATCAGGACCACTTG 581
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
 QY 1370 TGCCACCGACACCACTGGCAGCATGCTCAAGTACCTGGATGACGTGCGGCTCTCCTG 1429
 DB 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

QY 582 CACAGATACCCCTTACCAAAACACACAGATGATCTGACCTCCTTGAATAATACCTGGCCAA 641
 DB 1430 CAGCGTGGCTTCCACACCCATGCCAGAGAGCTTACTACCTGAACAAGCTGTCTCTCCAT 1489
 QY 642 CATCCGTTTGGATTCTGTTCTCAGGATGCAACAAGATTTGATGAGTCCAGGTTAGA 701
 DB 1490 CATGCTGGGCACACAGACCTGCTCCGGAGCGCTTACGCTGCTCAGTGCCTGGCTGA 1549
 QY 702 CACTGAAGTAGCCAACTTATCAGTATGATGAAAGAAATGAAGCTAGTAGACTCCAAGA 761
 DB 1550 CCTCAACGTCGGAACTCTTCCATGATCGTGAGAGAGATGAAGGAGTGGACACACAGA 1609
 QY 762 TGGTCAAGTCAACAAGATTTTACAAATCTACAAGTCTCCACCGGGCCCCAGGGTCCAA 821
 DB 1610 TGGAGAAATCTTTCGCAATGTCAACCATCTACGAGTGCCTCCCGCCCTCCAGGACCA 1669
 QY 822 AGGTACAGAGATCCCAAGGACCCCTGGCCCACTGGCAACAAGGAGACAAAGGAGA 881
 DB 1670 AGGATTCAAAGGAGATATGGCGTGAAGGGCTCTTGGCGGCAGAGGCCCGGAAAGGAGA 1729
 QY 882 GAAAGGGAGCGCTGGACCACTGGCCCTGCGGGTGAGAGAGGCCCAATTTGGACCACTGG 941
 DB 1730 CCCCGCATCTTGGGCCCTTGGGACCCCGAGGCTCTCAGGGGCAACCTTGGAGAGCCGG 1789
 QY 942 TCCCCCGGAGAGCGTGGCGCAAGGATCTAAAGGCTCCAGGGCCCCCAAGGCTCCCG 1001
 DB 1790 GCCTGTGGGAGAAAGGGGCCCTGTTGGCCTCGAGGTTTCCAGGCTCAAGGCTCAA 1849
 QY 1002 TGGTTCCTTGGGAAGCGGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGGCC 1061
 DB 1850 GGGCAGCTTTTGAACCTGGAGGGCCGAGAGAGCAGCCAGGCCCAAAAGGGGACATAGGGCC 1909
 QY 1062 ACCAGGCAAGAGGAGCTCCCGGCCCTCAGGGCCCTCCTGCTTCCAGGGACTTCAGGG 1121
 DB 1910 CCCAGGGCAGAAAGGGCCCCCGGGGTCTCCAGGGCCCTCAGGGCCCTCAGGGAAAACCGG 1969
 QY 1122 CACCGTTGGGAGCGCTGGGGTGCCTTGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACC 1181
 DB 1970 AATTGCAGGAAAGACAGGCTCACAGGCCAGCGGGGGCCATGGGGCTTAAGGGTGAACC 2029
 QY 1182 AGGATGCGAGGCCCAAGGGCCCCCGGGCCCTCTTGCCCTCATCAGGA 1230
 DB 2030 AGGGATCCAGGTTCCCTGGTCTCCCGGGGCTTCCAGGTCCACCAGGA 2078

Search completed: March 21, 2003, 02:18:38
 Job time : 362.008 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	121.6	7.2	810	1	US-08-642-255-60	Sequence 60, Appl	
2	118	7.0	756	1	US-08-642-255-50	Sequence 50, Appl	
3	113.6	6.7	3181	1	US-08-653-086-1	Sequence 1, Appl	
c	112.8	6.7	5062	1	US-08-253-155A-61	Sequence 61, Appl	
5	106	6.3	432	1	US-08-642-255-48	Sequence 48, Appl	
6	106	6.3	1608	3	US-09-029-348-19	Sequence 19, Appl	
7	105.4	6.2	503	1	US-08-023-155A-20	Sequence 20, Appl	
8	100	5.9	5102	1	US-08-494-168-1	Sequence 1, Appl	
9	99.6	5.9	1416	1	US-07-621-091G-1	Sequence 1, Appl	
10	99.6	5.9	1416	2	US-07-399-889-1	Sequence 1, Appl	
11	99.6	5.9	1416	3	US-09-167-364-1	Sequence 1, Appl	
12	99.6	5.9	1416	4	US-09-430-897-1	Sequence 1, Appl	
13	99.4	5.9	1868	1	US-08-392-367B-1	Sequence 1, Appl	
14	99.4	5.9	1868	3	US-08-893-467A-1	Sequence 1, Appl	
15	99.4	5.9	2543	1	US-08-555-669-11	Sequence 11, Appl	
16	99.4	5.9	2543	3	US-09-073-663-11	Sequence 11, Appl	
17	97.6	5.8	4359	4	US-09-484-970B-4	Sequence 4, Appl	
18	96.2	5.7	1572	4	US-09-297-269-39	Sequence 39, Appl	
19	95.2	5.6	1560	2	US-08-794-795-5	Sequence 5, Appl	
20	95.2	5.6	1560	4	US-09-249-200-5	Sequence 5, Appl	
21	95.2	5.6	1703	2	US-08-794-795-1	Sequence 1, Appl	
22	95.2	5.6	1703	4	US-09-249-200-1	Sequence 1, Appl	
23	89	5.3	1881	4	US-09-029-348-20	Sequence 20, Appl	
24	77	4.5	1560	4	US-09-453-702B-264	Sequence 264, App	
25	77	4.5	61663	4	US-09-453-702B-62	Sequence 62, Appl	
c	76.8	4.5	5552	3	US-08-155-888-1	Sequence 1, Appl	
27	74.6	4.4	38584	4	US-09-453-702B-50	Sequence 50, Appl	


```

RESULT 4
US-08-253-155A-61/c
: Sequence 61, Application US/08253155A
: Patent No. 5691147
: GENERAL INFORMATION:
: APPLICANT: Gyuris, Jeno
: APPLICANT: Draetta, Giulio
: TITLE OF INVENTION: CDK4 Binding
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII(text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/253,1
: FILING DATE: 02-JUN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709

```

```

RESULT 5
US-08-642-255-48
; Sequence 48, Application us/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```


US-09-029-348-19

Query Match 6.3%; Score 106; DB 4; Length 1608;
Best Local Similarity 52.8%; Pred. No. 1.3e-18;
Matches 229; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 796 GTGTCACCGGGCCAGGGGTCCAAGAGGTGACAGAGATCCCAGGGACCCCTGGCCCCA 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GGTGAACCTGGTCAAACTGGTCTGCAGTGCAGCTGGTCTCATATGGCCCGTGCGTCT 420

QY 856 ACTGGCAACAGGACAGAAGAGAGAGAGGGGAGCCTGACACCACCTGGGCCCTGCGGGT 915
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 GCTGGCAAAATGGAACCCGTGGTGAACCTGGTCTTCTGGTCTGTGGTCTGCTGGT 480

QY 916 GAGAGAGCCCAATTGGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAAAGGATCTAAA 975
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 GCTCTTGGCCCAAGAGTCTTAGTGGCCACAAGCATTCGTGGCGATAAGGAGAGCCC 540

QY 976 GGCTCCAGGGCCCCAAGGCTCCCGTGGTTCCCTCGGAAAGCCCGGCCCTCAGGGCCCC 1035
|| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GGTGAAGGGGCCAGAGGTCTTCTGGCTCAAGGGACACAATGGATTGCAAGGTGTG 600

QY 1036 AGTGGGACCCAGSCCCCCCGGGCCCAAGGAGGACATCCCCGGGCCCTCAGGGC 1095
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 CTGTGATTCGTGTGTACCATGGTGAAGTGCTCTCTGGCTCCGTGGGTCTGCTGGT 660

QY 1096 CTTCTGGCTTCCAGGACTTCAGGGCACCTTGGGGAGCCTGGGGTGCCTGGACCTCGG 1155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 CCTAGSGCCCTGCTGTCTCTTGGCCCTGCTGGAAAGATGGTCGACATGGACATCCT 720

QY 1156 GGACTGCCAGGTTCCTGGGGTACCAGGATGCCAGGCCCCCAAGGGCCCCCGGGCCCT 1215
|| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GSTACGGTTGGACTGCTGGCATTCGAGGGCCCTCAGGGTCAACCAAGGCCCTGCTGGCCCC 780

QY 1216 CTTGGCCCATCAGG 1229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 CCTGGTCCCCCTGG 794

RESULT 7
US-08-253-155A-20
; Sequence 20, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jenö
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MI-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:

```
; ; LENGTH: 503 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: cDNA
US-08-253-155A-20

Query Match
Best Local Similarity 6.2%; Score 105.4; DB 1; Length 503;
Matches 226; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 387 CAACGAGCAGACCTGAAGACCTGCAGACCTTACCAAGATGCGAGAAATAGAACACG 446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 CCACGAGAGACATGCATGATCTTCAGTACCATACCCACTACGCCAGAACCGCACTGT 74

QY 447 CATCAAGTTCAACCAACTGAGAGAGCTTCCAGCTCTTTGAGACGATATTGTGAACAT 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 GGAGAGTTTGATCTCTGGTAGGAGCAATGGCTTCTCAGAGATTGAAATTGGCAACCAT 134

QY 507 CATTAGCAATATCAGTTACACAGCCACACCTCGGAGCGCTGACCAATCTTAATGA 566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 CTTCAACCAATCAATGCCACCGACAAACACCGCACACATGCTCATGTACCTGGATGA 194

QY 567 AGTCAGACCACTTGACAGATACCCCTTACCAACACACAGATGATCTGACCTCTTGAA 626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 CGTGGCGCTCTCTGCACGCTGGGCTTCCACACCCATGCCGAGAGCTCTACTACCTGAA 254

QY 627 TAATACCTCGGCAACATCGTTGGATCTGTCTCTCAGATGCAACAAGATTGTAT 686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 CRAAGTCTGTCTCATCATGCTGGGACCAACAGACCTGCTCCGGAGCGCTTCAGCCTGCT 314

QY 687 GAGGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATATATGGAAGAATGAAGCT 746
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 CAGTGGCGGCTGGACCTCAAGCTCCGGAACCTCTCCATGATCGTGGAGGATGAAGG 374

QY 747 AGTAGACTCCAGCAGCTGCTAGCTCATCAAGATTTTACAAATCTACAAAGTCCACCGGG 806
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 AGGGACACACAGAAATGGGGAGATCCTTCGGAATGTAACATCTCTAGAGGTGCCCCCGC 434

QY 807 CCCCAG 813
    |||||
Db 435 CTCACG 441

RESULT 8
US-08-494-168-1
; Sequence 1, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
```

```
; ; NAME: SAXE, Bernhard D.
; ; REGISTRATION NUMBER: 28,665
; ; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (202)672-5300
; ; TELEFAX: (202)672-5399
; ; TELEX: 904136
; ; INFORMATION FOR SEQ ID NO: 1:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 5102 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: double
; ; TOPOLOGY: linear
; ; FEATURE:
; ; NAME/KEY: CDS
; ; LOCATION: join(2..82, 86..97, 101..4399, 4403..4420, 4424
; ; LOCATION: ..4465, 4469..4876, 4880..5101)
US-08-494-168-1

Query Match
Best Local Similarity 5.9%; Score 100; DB 1; Length 5102;
Matches 220; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 792 ACAAGTCCACCGGGCCCCCAGGGTCCAAAGAGGTGACAGAGGATCCAGGGACCCCTGG 851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2380 AAATGTTGCTCCGGGGGAAACAAGGCTTACAAGGATTAACAGGGGCACAAAGGATTTCTTGG 2439

QY 852 CCCAACTGCAACAGGGGACAGAAAGAGAGAGAGGGGAGCTGGACACCTGGCCCTGC 911
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2440 AGACTCTGGCGTTCCAGGACTCAAGGTTGTGCACGGGAGCTGGCTTACTAGGCCCA 2499

QY 912 GGGTGAGAGAGGCCCAATTTGGACAGCTGTCCTCCCGGAGAGCGCTGGGGCAAGGATC 971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2500 AGGTGAGCGGGGACGCTTGGGACACCCAGGACAGGTGGACAGCCAGGACCCAGGATC 2559

QY 972 TAAAGCTCCAGGGGCCCCAAAGGCTCCCGTGGTTTCCCTTGGGAAGCCCGGCCCTCAGG 1031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2560 TAGTGTCTCATGATGGCATCAAGGGCAATCTGGGCTCTCCAGGAGCACCAGGCTTCCAGG 2619

QY 1032 CCCCAGTGGGACCGAGGCCCCCGGCCCCCAGGCAAGAGGACTCCCGGCCCTCA 1091
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2620 CATCTCAGGACATCTCTGGAAGAAAGAACAGAGGCAAGAAAGTCTCTCTGGATCAAT 2679

QY 1092 GGGCCCTCTGCTTCCAGGGACTTCAGGGCACCTTGGGGAGCTTGGGTGCTGTGACC 1151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2680 TGTAAAGAAAGGCTCCAGGGCTAAAGGCTTCTCTGGAAATCCAGGCTTAGTAGACT 2739

QY 1152 TCGGGGACTGCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCCCAAGGCCCCCCCGG 1211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2740 GAAAGAGCCAGGCTCTCCAGGGTCTCGTGGTTCGCCAGGCTCTCTGGACCAAGG 2799

RESULT 9
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 542408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
```

Query Match 5.9%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

; FILE REFERENCE: 951263B

;; CURRENT APPLICATION NUMBER: US/09/167.364
;; CURRENT FILING DATE: 1998-10-07
;; EARLIER APPLICATION NUMBER: 08/399889
;; EARLIER FILING DATE: 1995-03-07
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Calf
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-167-364-1

Query Match 5.9%; Score 99.6; DB 3; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 796 GGTCCACCGGGCCCCAGAGGTCCAAGAGGTGACAGAGTCCAGGAGCCCTCGGCCCA 855
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 283 GGAGCACCAGGTGTCCCGGTGAGCCAGGGCCAGAGGTGATCCTGGATTCTATGATTT 342
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 856 ACTGCGACACAGGGACAGACAGAGAGAGAGAGGGGAGCCTGGACCACTGGCCCTCGGGGT 915
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 343 CCAGGCATGAAGAGGAAGAGGTAAATTCAGGATTTCCAGGACCACTGGACCTCCAGGG 402
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 916 GAGAGAGGCCCAATTGGACCACTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCTAAA 975
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 403 CAAAGTGACCAAAAGGACCACTGGATGATTCAGGATTTCCAGGACCACTGGACCTCCAGGG 462
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 976 GGTCTCCAGGGCCCCAAAGGCTCCGTTCCCTGGGAAGCCGGCCCTCAGGGCCCC 1035
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 463 TCCCTTCCAGGAAGCCAGGCCCACTGTGTTCCAGTGGAGAACAGGGATGCAAGGAA 522
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1036 AGTGGGACCAAGGCCCGCCCGGCGCCACAGCAAGAGAGGACTCCCGGCCCTCAGGGC 1095
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 523 CCGGGCCCCAGGACCACTGTGTTCCAGTGGAGAACAGGGATGCAAGGAA 582
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1096 CCTCTCGGCTTCCAGGACCTTCAGGGCACTTCAGGGCACTGGGGAGCCTGGGTGCTGACCTCGG 1155
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 583 GAGGATGTCTCCAGGAACTCTGGACCACTGGAGAAAAGGCAACAAAGGTTGTAAA 642
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1156 GACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCT 1215
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 643 GGAGAGCAAGGACCACTGGATCCGATGCTGCTGCGGCTTGAAGGGGAAACCTTGAGAC 702
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1216 CTGCGCCCATCAGGAGCGGTGG 1237
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 703 ACTGGACCACTGCAGCAGGGG 724
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 12
US-09-439-897-1
;; Sequence 1, Application US/09439897
;; Patent No. 6277558
;; GENERAL INFORMATION:
;; APPLICANT: Hudson, Billy G
;; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
;; FILE REFERENCE: 95-1263-C
;; CURRENT APPLICATION NUMBER: US/09/439,897
;; CURRENT FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 65
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-439-897-1

Query Match 5.9%; Score 99.6; DB 4; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 796 GGTCCACCGGGCCCCAGAGGTCCAAGAGGTGACAGAGTCCAGGAGCCCTCGGCCCA 855
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 283 GGAGCACCAGGTGTCCCGGTGAGCCAGGGCCAGAGGTGATCCTGGATTCTATGATTT 342
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 856 ACTGCGCAACAGGGACAGACAGAGAGAGAGGGGAGCCTGGACCACTGGCCCTCGGGGT 915
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 343 CCAGGCATGAAGAGGAAGAGGTAAATTCAGGATTTCCAGGACCACTGGACCTCCAGGG 402
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 916 GAGAGAGGCCCAATTGGACCACTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCTAAA 975
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 403 CAAAGTGACCAAAAGGACCACTGGATGATTCAGGATTTCCAGGACCACTGGACCTCCAGGG 462
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 976 GGTCTCCAGGGCCCCAAAGGCTCCGTTCCCTGGGAAGCCGGCCCTCAGGGCCCC 1035
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 463 TCCCTTCCAGGAAGCCAGGCCCACTGTGTTCCAGTGGAGAACAGGGATGCAAGGAA 522
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1036 AGTGGGACCAAGGCCCGCCCGGCGCCACAGCAAGAGAGGACTCCCGGCCCTCAGGGC 1095
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 523 CCGGGCCCCAGGACCACTGTGTTCCAGTGGAGAACAGGGATGCAAGGAA 582
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1096 CCTCTCGGCTTCCAGGACCTTCAGGGCACTTCAGGGCACTGGGGAGCCTGGGTGCTGACCTCGG 1155
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 583 GAGGATGTCTCCAGGAACTCTGGACCACTGGAGAAAAGGCAACAAAGGTTGTAAA 642
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1156 GACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCT 1215
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 643 GGAGAGCAAGGACCACTGGATCCGATGCTGCTGCGGCTTGAAGGGGAAACCTTGAGAC 702
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 1216 CTGCGCCCATCAGGAGCGGTGG 1237
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 703 ACTGGACCACTGCAGCAGGGG 724
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 13
US-08-392-367B-1
;; Sequence 1, Application US/08392367B
;; Patent No. 5691197
;; GENERAL INFORMATION:
;; APPLICANT: Tryggvason, Karl
;; APPLICANT: Elomaa, Outi
;; APPLICANT: Kangas, Maarit
;; TITLE OF INVENTION: An Insolated DNA Sequence For a
;; Patent No. 5691197
;; TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
;; TITLE OF INVENTION: a Collagenous Domain and the
;; TITLE OF INVENTION: Polypeptide Chain Encoded by
;; TITLE OF INVENTION: such a Sequence
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
;; ADDRESSEE: Minnich & McKee
;; STREET: 1100 Superior Avenue
;; STREET: Suite 700
;; CITY: Cleveland
;; STATE: Ohio
;; COUNTRY: U.S.A.
;; ZIP: 44114-2518
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch,
;; MEDIUM TYPE: 720 Kb storable
;; COMPUTER: IBM PS/2, Model 35 SX
;; OPERATING SYSTEM: DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/392,367B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Minnich, Richard J.

REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 2 009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1868 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Nucleotide-genomic DNA
HYPOTHETICAL: No. 5691197 relevant
ANTI-SENSE: No. 5691197 relevant
US-08-392-367B-1

Query Match 5.9%; Score 99.4; DB 1; Length 1868;
Best Local Similarity 51.5%; Pred. No. 7.6e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 793 CAAGGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACAGGACCCCTGCG 852
Db 721 CCAGGTGTCCAGGACCCAGGGCCCCACAGGAGCAAGGAGAGGAGCGACCTCCAGGGA 780
QY 853 CCAACTGGCAACAAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 912
Db 781 CTTACGGTCCACCGAGGAAGCAAGGAGCAACTGTGCTCCAGGACCTCGAGGAGAGAAG 840
QY 913 GGTGAGAGAGGCCAATTTGGACAGCTGTGTCCTCCCGGAGAGAGGTGGCGCAAGGATCT 972
Db 841 GGCAGCAAAAGTGACATAGTCTCACTGGCCCCCAAGGGGGAACATGGCACCAGGGAGAG 900
QY 973 AAAGGTCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTTGGGAGCGCCGCTCAGGCG 1032
Db 901 AAAGGGACCTAGGCTTCCAGGAACAAAGGGAGCATGGGATGAAGGAGAGACAGCGGG 960
QY 1033 CCAAGTGGGACCCAGGCCCCCGGCCCCACAGGCAAAAGAGGAGTCCCGGCCCCCTCAG 1092
Db 961 CCATGGGTCCCTGGAGTCAGGAGGTAAAGGTGATGCTGGAACAAACAGGCTTACCA 1020
QY 1093 GGCCTCTCTGGTTCAGGAGACTTCAGGACACCGTTGGGGAGCGCTGGGTGCTGACCT 1152
Db 1021 GGTTCGGCTGGATCTCCAGGAGTCAAGGTGACCAAGGAAACCTGGAGTGCAGGGTGT 1080
QY 1153 CGGGAGCTCCAGGCTGCTGGGGTACCAAGGATCCAGGCCCCCAAGGGCCCCCGCGC 1212
Db 1081 CCAGGCCCTCAAGGTCCACAGGACTTTTCAGGTGCCAAGGTGAGCCAGGACGACTGGT 1140
QY 1213 CTTCTGGGCCATCAGGAGCGGTGG 1237
Db 1141 CTTCTGGGCCAGCAGGACCCCGG 1165

RESULT 14

US-08-893-467A-1

Sequence 1, Application US/08893467A

Patent No. 6063901

GENERAL INFORMATION:

APPLICANT: Tryggvason, Karl

APPLICANT: Elomaa, Outi

APPLICANT: Kangas, Maarit

TITLE OF INVENTION: An Isolated DNA Sequence For a

Patent No. 6063901

TITLE OF INVENTION: No. 6063901el Macrophage Receptor with

TITLE OF INVENTION: a Collagenous Domain and the

TITLE OF INVENTION: Polypeptide Chain Encoded by

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fay, Sharpe, Beall, Fagan,

ADDRESSEE: Minnich & McKee

STREET: 1100 Superior Avenue

STREET: Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: U.S.A.
ZIP: 44114-2518
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch,
MEDIUM TYPE: 720 Kb storable
COMPUTER: IBM PS/2, Model 35 SX
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,467A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard J.
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 2 009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1868 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Nucleotide-genomic DNA
HYPOTHETICAL: No. 6063901 relevant
ANTI-SENSE: No. 6063901 relevant
US-08-893-467A-1

Query Match 5.9%; Score 99.4; DB 3; Length 1868;

Best Local Similarity 51.5%; Pred. No. 7.6e-17;

Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 793 CAAGGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACAGGACCCCTGCG 852
Db 721 CCAGGTGTCCAGGACCCAGGGCCCCACAGGAGCAAGGAGAGGAGCGACCTCCAGGGA 780
QY 853 CCAACTGGCAACAAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 912
Db 781 CTTACGGTCCACCGAGGAAGCAAGGAGCAACTGTGCTCCAGGACCTCGAGGAGAGAAG 840
QY 913 GGTGAGAGAGGCCAATTTGGACAGCTGTGTCCTCCCGGAGAGAGGTGGCGCAAGGATCT 972
Db 841 GGCAGCAAAAGTGACATAGTCTCACTGGCCCCCAAGGGGGAACATGGCACCAGGGAGAG 900
QY 973 AAAGGTCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTTGGGAGCGCCGCTCAGGCG 1032
Db 901 AAAGGGACCTAGGCTTCCAGGAACAAAGGGAGCATGGGATGAAGGAGAGACAGCGGG 960
QY 1033 CCAAGTGGGACCCAGGCCCCCGGCCCCACAGGCAAAAGAGGAGTCCCGGCCCCCTCAG 1092
Db 961 CCATGGGTCCCTGGAGTCCAGGAGGTAAAGGTGATGCTGGAACAAACAGGCTTACCA 1020
QY 1093 GGCCTCTCTGGTTCAGGAGACTTCAGGACACCGTTGGGGAGCGCTGGGTGCTGACCT 1152
Db 1021 GGTTCGGCTGGATCTCCAGGAGTCAAGGTGACCAAGGAAACCTGGAGTGCAGGGTGT 1080
QY 1153 CGGGAGCTCCAGGCTGCTGGGGTACCAAGGATCCAGGCCCCCAAGGGCCCCCGCGC 1212
Db 1081 CCAGGCCCTCAAGGTCCACAGGACTTTTCAGGTGCCAAGGTGAGCCAGGACGACTGGT 1140
QY 1213 CTTCTGGGCCATCAGGAGCGGTGG 1237
Db 1141 CTTCTGGGCCAGCAGGACCCCGG 1165

RESULT 15

US-08-553-669-11

```

Db      944   CCGGGATGCCAGCAAGGACGCCAGAAATGGCGTCCCAAGGACTCGATGGCCAGAGGGA  1003
Qy      1213   CCTCCTGGCCCCATCAGAGCGGTGG  1237
          |||||  |||||  ||
Db      1004   GAGCTGGTCGCAACGGTGCTCCGG  1028

```

Search completed: March 21, 2003, 08:25:28
Job time : 79.7466 secs

[illegible]

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 269.408 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712a-1_COPY_670_2024

Perfect score: 1355

Sequence: 1 atgcaacaagattgatgag.....agaaaagggttatcatcccg 1355

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1355	100.0	2024	21	AAA07697 Human collectin en
2	1355	100.0	2628	22	AAH43036 Nucleotide sequenc
3	1337.4	98.7	2005	24	ABL95574 Human angio genesis
4	1337.4	98.7	2005	24	ABL88085 Human PRO7223 cDNA
5	1337.4	98.7	2318	22	AAI58842 Human polynucleoti
6	1337.4	98.7	2641	22	AAC66903 Human EXMAD-14 cod
7	1335.8	98.6	2930	20	AAV55746 Human secreted pro
8	1328.8	98.1	2181	22	AAI60628 Human polynucleoti
9	1323.8	97.7	2929	24	ABQ92072 Human polynucleoti

10	1029	75.9	2262	24	ABA97932 Human scavenger re
11	895.2	66.1	2637	22	AAH43037 Nucleotide sequenc
12	883.8	65.2	1521	23	AA571133 DNA encoding novel
13	743.8	54.9	2256	22	AAH43054 Nucleotide sequenc
14	413	30.5	493	23	AA571131 DNA encoding novel
15	332	24.5	5041	22	ABAL7153 Human nervous syst
16	249	18.4	415	24	ABL66665 Lung cancer relate
17	157.6	11.6	3685	20	AA527858 Human CSR1 protein
18	157.6	11.6	3810	20	AA527856 Human CSR1 protein
19	150.2	11.1	873	23	AA571130 DNA encoding novel
20	150.2	11.1	1062	23	AA567442 DNA encoding novel
21	146.8	10.8	1329	23	AA571134 DNA encoding novel
22	121.6	9.0	756	14	AAQ43034 Collagen-like poly
23	121.6	9.0	756	17	AA116768 Collagen-like poly
24	121.4	9.0	5676	15	AA064556 Human collagen (TY
25	121.4	9.0	8284	22	AA126526 Human breast cance
26	121.4	9.0	8284	22	AA126553 Human breast cance
27	121.4	9.0	8284	22	AA126600 Human breast cance
28	118.4	8.7	6109	23	AAH78667 Murine Col5a3 cDNA
29	118	8.7	756	14	AAQ43032 Collagen-like poly
30	118	8.7	756	17	AA116766 Collagen-like poly
31	117	8.6	9287	24	ABK64501 Human benign prost
32	115.8	8.5	4748	22	AA066573 Bovine alpha1(I) c
33	115	8.5	4270	24	AB199885 Mouse ischaemic co
34	114.8	8.5	4816	24	ABK84041 Human cDNA differe
35	114.8	8.5	5058	23	AA569026 DNA encoding novel
36	114.8	8.5	6358	22	AAH98286 Human EST-derived
37	114.8	8.5	6691	23	AA579806 DNA encoding novel
38	114	8.4	5060	24	AAH92176 Prostate cancer-as
39	113.6	8.4	3181	19	AAV59338 Nucleotide sequenc
40	113.4	8.4	4821	23	AA586866 DNA encoding novel
41	113.4	8.4	6200	23	AAH78668 Human COL5A3 cDNA
42	113.4	8.4	6368	24	ABN59647 Novel human coding
43	113.2	8.4	2319	13	AAQ24551 DNA encoding human
44	113.2	8.4	2319	18	AAH47812 Human type IV coll
45	113.2	8.4	3898	22	AAH76377 Murine adipocytes-

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.

XX AC AAA07697;

XX DF 20-JUN-2000 (first entry)

XX DE Human collectin encoding DNA.

XX DE Collectin; human; antibacterial; antiviral; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 55..1698

XX FT /*tag= a

XX FT /product= "collectin"

XX FT misc_feature 1..738

XX FT /*tag= b

XX FT misc_feature 55..738

XX FT /*tag= c

XX FT misc_feature 79..738

XX FT /*tag= d

XX FT misc_feature 325..738

XX FT /*tag= e

XX FT misc_feature 358..738

XX FT /*tag= f

FT misc_feature /note= "fragment specifically claimed in claim 8"
670..1695
FT /*tag= g
FT /note= "fragment specifically claimed in claim 5"
685..738
FT /*tag= h
FT /note= "fragment specifically claimed in claim 7"
730..738
FT /*tag= i
FT /note= "fragment specifically claimed in claim 7"
739..1695
FT /*tag= j
FT /note= "fragment specifically claimed in claim 6"
1696..2024
FT /*tag= k
FT /note= "fragment specifically claimed in claim 9"
XX
XX WO200011161-A1.
XX
XX 02-MAR-2000.
PD
XX 24-AUG-1999; 99WO-JP04552.
XX
XX 24-AUG-1998; 98JP-0237611.
PR
XX (FUSO) FUSO PHARM IND LTD.
PA
XX Wakamiya N;
XX
XX WPI; 2000-224696/19.
DR P-PSDB; AAY77985.
XX
XX New collectin of human origin having antibacterial and antiviral
PT activity, and gene encoding it useful for production of transgenic
PT animals and of antibodies for screening potential drug molecules -
XX
XX Claim 5-9; Page 87-90; 106pp; Japanese.
PS
XX The invention relates to polynucleotides encoding a new collectin of
CC human origin. The collectin can be used as an antibacterial and antiviral
CC agent and for screening potential drug molecules. The new collectin can
CC be produced by standard recombinant methodology. The present sequence
CC represents a DNA encoding the human collectin.
XX
SQ Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;
Query Match 100.0%; Score 1355; DB 21; Length 2024;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 670 ATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 729
Qy 61 ATGGAAGAAATGACGTAGTAGCTCCAAAGCATGGTTCAGCTCATCAAGAAATTTACAAATA 120
Db 730 ATGGAAGAAATGACGTAGTAGCTCCAAAGCATGGTTCAGCTCATCAAGAAATTTACAAATA 789
Qy 121 CTCAAGGTCTCACCGGGCCCGAGGGTCCAAAGAGTGCACAGGATCCAGGGACCCCT 180
Db 790 CTCAAGGTCTCACCGGGCCCGAGGGTCCAAAGAGTGCACAGGATCCAGGGACCCCT 849
Qy 181 GCGCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 240
Db 850 GCGCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 909
Qy 241 GCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCGCCCGAGAGCGGTGGCGGCAAAAGGA 300
Db 910 GCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCGCCCGAGAGCGGTGGCGGCAAAAGGA 969
Qy 301 TCTAAGGCTCCAGGGCCCAAGAGTCCCGTGGTTCCTGGAGAGCCCGCCCTCAG 360
Db 970 TCTAAGGCTCCAGGGCCCAAGAGTCCCGTGGTTCCTGGAGAGCCCGCCCTCAG 1029

RESULT 2

AAH43036

ID AAH43036 standard; DNA; 2628 BP.

Qy 361 GGGCCCAAGTGGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGAGTCCCCGGCCCT 420
Db 1030 GGGCCCAAGTGGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGAGTCCCCGGCCCT 1089
Qy 421 CAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCCTGGGGAGCCCTGGGGTGCCTGGA 480
Db 1090 CAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCCTGGGGAGCCCTGGGGTGCCTGGA 1149
Qy 481 CTGGGGAGTGGCCAGGCTTGGCTGGGTACACAGGCATGCCAGGCCCAAGGCCCGCCCC 540
Db 1150 CTGGGGAGTGGCCAGGCTTGGCTGGGTACACAGGCATGCCAGGCCCAAGGCCCGCCCC 1209
Qy 541 GGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCAAACCCG 600
Db 1210 GGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCAAACCCG 1269
Qy 601 GCACGGAGGACAAATGGCTGCCCGCCCTACCTGGAGAACTTCACAGACAAATGCTACTAT 660
Db 1270 GCACGGAGGACAAATGGCTGCCCGCCCTACCTGGAGAACTTCACAGACAAATGCTACTAT 1329
Qy 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAAGCTTTCA 720
Db 1330 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAAGCTTTCA 1389
Qy 721 CATCTTGTTCATAAACAACACTAGAGAGGACAGCAATGGATAAAAAACAGATGGTAGGG 780
Db 1390 CATCTTGTTCATAAACAACACTAGAGAGGACAGCAATGGATAAAAAACAGATGGTAGGG 1449
Qy 781 AGAGAGCCCACTGGATGGCTCACAGCTCACAGCTCAGAGCGTGAAATGAATGAAGTGGCTG 840
Db 1450 AGAGAGCCCACTGGATGGCTCACAGCTCACAGCTCAGAGCGTGAAATGAATGAAGTGGCTG 1509
Qy 841 GATGGACATCTCAGACTACAAAATTTGAAAGCTGGACAGCGGATAACTGGGGTCTAT 900
Db 1510 GATGGACATCTCAGACTACAAAATTTGAAAGCTGGACAGCGGATAACTGGGGTCTAT 1569
Qy 901 GGCCATGGCCAGGAGAGACTGTCTGGGTGATTTATGTGGGCACTGGAACGATTTTC 960
Db 1570 GGCCATGGCCAGGAGAGACTGTCTGGGTGATTTATGTGGGCACTGGAACGATTTTC 1629
Qy 961 CAATGTGAAGACGTCAATAACTTCAATTTGCGAAAAGACAGGGAGACAGTACTGTCATCT 1020
Db 1630 CAATGTGAAGACGTCAATAACTTCAATTTGCGAAAAGACAGGGAGACAGTACTGTCATCT 1689
Qy 1021 GCATTATAACGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGGCAAGG 1080
Db 1690 GCATTATAACGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGGCAAGG 1749
Qy 1081 ACACCTCTTTCTAATTCATCACCCTTCTCATCAGATTGAAAAAAGCACTGAAAA 1140
Db 1750 ACACCTCTTTCTAATTCATCACCCTTCTCATCAGATTGAAAAAAGCACTGAAAA 1809
Qy 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTTTTACCATCCGTCATTTACCAAGAC 1200
Db 1810 CCAATTACTGAAAAAATTTGACAGCTAGTGTTTTACCATCCGTCATTTACCAAGAC 1869
Qy 1201 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTTGTGACACATGACTGAA 1260
Db 1870 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTTGTGACACATGACTGAA 1929
Qy 1261 TCACATAGATTCTCCTCCGTGAGTAACCGTGCAGATTATACAAATTTATGCTTTCCAAAGTA 1320
Db 1930 TCACATAGATTCTCCTCCGTGAGTAACCGTGCAGATTATACAAATTTATGCTTTCCAAAGTA 1989
Qy 1321 TGGAACTCTCAATCAGAAAAAGGTTATCATCCCG 1355
Db 1990 TGGAACTCTCAATCAGAAAAAGGTTATCATCCCG 2024

Db	2594	TGGAACACTCCAATCAGAAAAAGGTTATCATCCCG	2628
RESULT 3			
ABL95574			
ID	ABL95574	standard; cDNA; 2005 BP.	
XX	AC	ABL95574;	
XX	DT	19-JUL-2002 (first entry)	
XX	DE	Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.	
XX	KW	Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;	
KW	KW	atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;	
KW	KW	cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;	
XX	KW	antiarteriosclerotic; gene; ss.	
OS	XX	Homo sapiens.	
XX	PN	WO200208284-A2.	
XX	PD	31-JAN-2002.	
XX	PF	09-JUL-2001; 2001WO-US21735.	
XX	PR	20-JUL-2000; 2000US-219556P.	
PR	PR	25-JUL-2000; 2000US-220624P.	
PR	PR	25-JUL-2000; 2000US-220664P.	
PR	PR	28-JUL-2000; 2000WO-US20710.	
PR	PR	02-AUG-2000; 2000US-222695P.	
PR	PR	17-AUG-2000; 2000US-0643657.	
PR	PR	23-AUG-2000; 2000WO-US23328.	
PR	PR	24-AUG-2000; 2000WO-US23328.	
PR	PR	07-SEP-2000; 2000US-230978P.	
PR	PR	15-SEP-2000; 2000US-000000P.	
PR	PR	18-SEP-2000; 2000US-0664610.	
PR	PR	18-SEP-2000; 2000US-0665350.	
PR	PR	24-OCT-2000; 2000US-242922P.	
PR	PR	08-NOV-2000; 2000US-0709238.	
PR	PR	08-NOV-2000; 2000WO-US30952.	
PR	PR	10-NOV-2000; 2000WO-US30873.	
PR	PR	01-DEC-2000; 2000US-0522678.	
PR	PR	20-DEC-2000; 2000US-0747259.	
PR	PR	22-DEC-2000; 2000WO-US34956.	
PR	PR	22-DEC-2001; 2001US-0767609.	
PR	PR	28-FEB-2001; 2001US-0796498.	
PR	PR	28-FEB-2001; 2001WO-US06520.	
PR	PR	01-MAR-2001; 2001WO-US06666.	
PR	PR	09-MAR-2001; 2001US-0802708.	
PR	PR	14-MAR-2001; 2001US-0808689.	
PR	PR	22-MAR-2001; 2001US-0816744.	
PR	PR	05-APR-2001; 2001US-0828366.	
PR	PR	10-MAY-2001; 2001US-0854208.	
PR	PR	20-MAY-2001; 2001US-0854280.	
PR	PR	25-MAY-2001; 2001US-0866028.	
PR	PR	25-MAY-2001; 2001WO-US17092.	
PR	PR	30-MAY-2001; 2001US-0870574.	
PR	PR	30-MAY-2001; 2001WO-US17443.	
PR	PR	01-JUN-2001; 2001WO-US17800.	
PR	PR	20-JUN-2001; 2001WO-US19692.	
PR	PR	28-JUN-2001; 2001WO-US00000.	
PA	(GETH)	GENENTECH INC.	
PA	(BAKE/)	BAKER K P.	
PA	(FERR/)	FERRARA N.	
PA	(GERB/)	GERBER H.	
PA	(GERR/)	GERRITSEN M E.	
PA	(GODD/)	GODDARD A.	
PA	(GODO/)	GODOWSKI P J.	
PA	(GURN/)	GURNEY A L.	
PA	(HILL/)	HILLAN K J.	
PA	(MARS/)	MARSTERS S A.	
PA	(PANJ/)	PAN J.	
PA	(PAON/)	PAONI N F.	
PA	(STEP/)	STEPHAN J F.	
PA	(WATA/)	WATANABE C K.	
PA	(WILL/)	WILLIAMS P M.	
PA	(WOOD/)	WOOD W I.	
XX	Baker KP,	Ferrara N,	Gerber H,
PI	Godowski PJ,	Gurney AL,	Hillan KJ,
PI	Stephan JF,	Watanabe CK,	Williams PM,
XX	WPI; 2002-171999/22.		
DR	P-PSDB; ABB95436.		
XX	One hundred and eighty seven nucleic acids encoding PRO polypeptides,		
PT	useful in diagnosis and treatment of cardiovascular (e.g. myocardial		
PT	infarction), endothelial or angiogenic disorders in a mammal -		
XX	Claim 1; Fig 27; 567pp; English.		
XX	The present invention provides the protein and coding sequences of human		
CC	PRO proteins. These are useful for treating or diagnosing a		
CC	cardiovascular, endothelial or angiogenic disorder, including cardiac		
CC	hypertrophy, trauma, cancer, age-related macular degeneration,		
CC	atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,		
CC	angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour		
CC	angiogenesis (such as breast carcinoma and liver carcinoma) and wound		
CC	healing. The present sequence is a coding sequence of the invention.		
XX	Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;		
SQ	Query Match 98.7%; Score 1337.4; DB 24; Length 2005;		
	Best Local Similarity 99.9%; Pred No. 0;		
	Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;		
Qy	1	ATGCAACAAGATTTCATGAGTTCGAGGTAGACACTGAAGTAGCCAACTATACGTGATT	60
Db	478	ATGCAACAAGATTTCATGAGTTCGAGGTAGACACTGAAGTAGCCAACTATACGTGATT	537
Qy	61	ATGGAAGAATCAAGCTAGTACCTCCAAAGCATGTCTAGCTCATCAAGATTTTACAATA	120
Db	538	ATGGAAGAATCAAGCTAGTACCTCCAAAGCATGTCTAGCTCATCAAGATTTTACAATA	597
Qy	121	CTACAAGGTCCACCGGGCCCAAGAGGTCCAAAGAGGTGACAGAGGATCCAGGACCCCT	180
Db	598	CTACAAGGTCCACCGGGCCCAAGAGGTCCAAAGAGGTGACAGAGGATCCAGGACCCCT	657
Qy	181	GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGAGGGAGCCTGGACCTGGCCCT	240
Db	658	GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGAGGGAGCCTGGACCTGGCCCT	717
Qy	241	CGGGGTGAGAGAGGCCCAATTGGACCACTGGTCCCTCCCGGAGAGCGTGGCGGCAAGA	300
Db	718	CGGGGTGAGAGAGGCCCAATTGGACCACTGGTCCCTCCCGGAGAGCGTGGCGGCAAGA	777
Qy	301	TCTAAAGGCTCCAGGGCCCAAGAGGTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG	360
Db	778	TCTAAAGGCTCCAGGGCCCAAGAGGTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG	837
Qy	361	GGCCCCACTGGGGACCCAGGCCCCCGGGCCCAACAGGATCCAGGAGAGGACTTCCCGGCCCT	420
Db	838	GGCCCCACTGGGGACCCAGGCCCCCGGGCCCAACAGGATCCAGGAGAGGACTTCCCGGCCCT	897
Qy	421	CAGGGCCCTCTGGCTTCAGGGACTTCAGGGACCGTTGGGGAGCCTGGGGTGCTGGA	480
Db	898	CAGGGCCCTCTGGCTTCAGGGACTTCAGGGACCGTTGGGGAGCCTGGGGTGCTGGA	957
Qy	481	CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCAAGGCCCGCCCC	540
Db	958	CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCAAGGCCCGCCCC	1017
Qy	541	GGCCCTCTGGCCCATCATCAGGAGCGGTGGTGGCCCTTGCAAGATGAGCCCAACCCCG	600

CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
XX probes used in the exemplification of the present invention.

SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Query Match 98.7%; Score 1337.4; DB 24; Length 2005;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCAACAGATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB 478 ATGCAACAGATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 537
QY 61 ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGGTTCAGCTCATCAAGAATTTTACAATA 120
DB 538 ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGGTTCAGCTCATCAAGAATTTTACAATA 597
QY 121 CTAAAGGTTCACCGGGCCCCAGGGTCCAAAGGTGACAGAGGTGCCAGGACCCGCT 180
DB 598 CTAAAGGTTCACCGGGCCCCAGGGTCCAAAGGTGACAGAGGTGCCAGGACCCGCT 657
QY 181 GCGCCAACTGGCAACAGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTGGCCCT 240
DB 658 GCGCCAACTGGCAACAGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTGGCCCT 717
QY 241 GCGGGTAGAGAGGCCCCAATTGGACAGCTGGTCCCGCCCGGAGAGCGTGGCGGCAAGGA 300
DB 718 GCGGGTAGAGAGGCCCCAATTGGACAGCTGGTCCCGCCCGGAGAGCGTGGCGGCAAGGA 777
QY 301 TCTAAAGGTCCTCAGGGCCCCAAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG 360
DB 778 TCTAAAGGTCCTCAGGGCCCCAAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG 837
QY 361 GCGCCCACTGGGAGCCAGGCCCCCGGCCCCCAGGAGGAGGAGGACTCCCGGCCCT 420
DB 838 GCGCCCACTGGGAGCCAGGCCCCCGGCCCCCAGGAGGAGGAGGACTCCCGGCCCT 897
QY 421 CAGGGCCCTCCTGGGTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGGTGGCTGGA 480
DB 898 CAGGGCCCTCCTGGGTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGGTGGCTGGA 957
QY 481 CTTGGGGAGTCCAGGCTTGGGTACCAAGCATGCCAGGCCCCCAAGGCCCCCCCC 540
DB 958 CTTGGGGAGTCCAGGCTTGGGTACCAAGCATGCCAGGCCCCCAAGGCCCCCCCC 1017
QY 541 GCGCCCTCCTGGGCCCTCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCAACCCCG 600
DB 1018 GCGCCCTCCTGGGCCCTCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCAACCCCG 1077
QY 601 GCACGGAGGACAATGGGCTGCGCCGCTCACTGGAGAACTTCACAGACAAATGCTACTAT 660
DB 1078 GCACGGAGGACAATGGGCTGCGCCGCTCACTGGAGAACTTCACAGACAAATGCTACTAT 1137
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGTGAAGCTTTCTGCAAGACAAGTCTTCA 720
DB 1138 TTTTCAGTTGAGAAAGAAATTTTGGAGTGAAGCTTTCTGCAAGACAAGTCTTCA 1197
QY 721 CATCTGTTGTTTCAATAACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 780
DB 1198 CATCTGTTGTTTCAATAACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 1257
QY 781 AGAGAGAGCCACTGGATCGGCCCTCAGACTCAGAGCGGTGAAAATGAATGGAAGTGGCTG 840
DB 1258 AGAGAGAGCCACTGGATCGGCCCTCAGACTCAGAGCGGTGAAAATGAATGGAAGTGGCTG 1317
QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCCGATACTGGGGTCAAT 900
DB 1318 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCCGATACTGGGGTCAAT 1377
QY 901 GGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAACGATTTC 960
DB 1378 GGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAACGATTTC 1437

QY 961 CAATGTGAAGAGCGTCAATAACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 1020
DB 1438 CAATGTGAAGAGCGTCAATAACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 1497
QY 1021 GCATTATACCGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGAGGCAAGG 1080
DB 1498 GCATTATACCGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGAGGCAAGG 1557
QY 1081 ACACATCCCTTTCTAATTTGATCATCACTTCTCATCAGATTGAAAAAAGAACGACTGAAAA 1140
DB 1558 ACACATCCCTTTCTAATTTGATCATCACTTCTCATCAGATTGAAAAAAGAACGACTGAAAA 1616
QY 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
DB 1617 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1676
QY 1201 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA 1260
DB 1677 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA 1736
QY 1261 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTATACAAATTTATGCTTCCAAAGTA 1320
DB 1737 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTATACAAATTTATGCTTCCAAAGTA 1796
QY 1321 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1351
DB 1797 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1827
RESULT 5
AA158842
ID AA158842 standard; cDNA; 2318 BP.
XX AC AA158842;
XX DF 22-OCT-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 1045.
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM39686.
PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -
PS Claim 1; SEQ ID NO 1045; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Query Match 98.7%; Score 1337.4; DB 22; Length 2318;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCAACAGATTTCATGAGTGTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB 663 ATGCAACAGATTTCATGAGTGTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 722

QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGTGCTAGCTCATCAAGAATTTTACAATA 120
DB 723 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGTGCTAGCTCATCAAGAATTTTACAATA 782

QY 121 CTACAGGTCCACGGGCCCGCCAGGGTCCAGAGGTGACAGAGATCCAGGGACCCCT 180
DB 783 CTACAGGTCCACGGGCCCGCCAGGGTCCAGAGGTGACAGAGATCCAGGGACCCCT 842

QY 181 GSCCAACTGGCAACAGGGACAGAAAGGAGAGAGGGGAGCCCTGGACACTGGCCCT 240
DB 843 GSCCAACTGGCAACAGGGACAGAAAGGAGAGAGGGGAGCCCTGGACACTGGCCCT 902

QY 241 GCGGTGAGAGAGGCCCAATTTGACACAGTGTGCCCGGAGAGCGTGGCGGCAAGGA 300
DB 903 GCGGTGAGAGAGGCCCAATTTGACACAGTGTGCCCGGAGAGCGTGGCGGCAAGGA 962

QY 301 TCTAAGGCTCCAGGGGCCCAAGAGGCTCCCGTGGTCCCTGGGAGCCCGGCCCTCAG 360
DB 963 TCTAAGGCTCCAGGGGCCCAAGAGGCTCCCGTGGTCCCTGGGAGCCCGGCCCTCAG 1022

QY 361 GSCCCAGTGGGACCCAGGCCCGCCCGGCCACAGGCAAGAGGGACTCCCGGCCCT 420
DB 1023 GSCCCAGTGGGACCCAGGCCCGCCCGGCCACAGGCAAGAGGGACTCCCGGCCCT 1082

QY 421 CAGGCCCTCTCGGTTCCAGGGACTTCAGGGCACCGTGGGGAGCCCTGGGTGCTGGA 480
DB 1083 CAGGCCCTCTCGGTTCCAGGGACTTCAGGGCACCGTGGGGAGCCCTGGGTGCTGGA 1142

QY 481 CCTCGGGACTCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 540
DB 1143 CCTCGGGACTCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 1202

QY 541 GSCCCTCTCGGCCATCAGGAGCGTGTGCCCTGGCCCTGCAGAAATGAGCAACCCCG 600
DB 1203 GSCCCTCTCGGCCATCAGGAGCGTGTGCCCTGGCCCTGCAGAAATGAGCAACCCCG 1262

QY 601 GCACCGGAGCAATGGTGGCCCGCTCAGTGAAGAATTTACAGACAAAATGCTACTAT 660
DB 1263 GCACCGGAGCAATGGTGGCCCGCTCAGTGAAGAATTTACAGACAAAATGCTACTAT 1322

QY 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAGTCTTCA 720
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Db 1323 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAGTCTTCA 1382
QY 721 CATCTTGTGTTTTCATAAAACACTAGAGAGCAACAGCAATGCGATAAAAAACAGATGTAGG 780
Db 1383 CATCTTGTGTTTTCATAAAACACTAGAGAGCAACAGCAATGCGATAAAAAACAGATGTAGG 1442
QY 781 AGAGAGAGCCACTGATGGCCCTCAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 840
Db 1443 AGAGAGAGCCACTGATGGCCCTCAGACTCAGAGCGTGAAACGAATGGAAGTGGCTG 1502
QY 841 GATGGACATCTCCAGACTTACAAAAATTTGGAAGCTGGACACCGGATCAACTGGGCTCAT 900
Db 1503 GATGGACATCTCCAGACTTACAAAAATTTGGAAGCTGGACACCGGATCAACTGGGCTCAT 1562
QY 901 GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGGAGTGGACGATTTTC 960
Db 1563 GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGGAGTGGACGATTTTC 1622
QY 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACACTACTGTCTCATCT 1020
Db 1623 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACACTACTGTCTCATCT 1682
QY 1021 GCATTATACGAGCTGTGATGGATCACATGAGCAAAATTTTCAGCTCTCAAAGGCCAAAG 1080
Db 1683 GCATTATACGAGCTGTGATGGATCACATGAGCAAAATTTTCAGCTCTCAAAGGCCAAAG 1742
QY 1081 ACACCTCCTTTCTAAATTGCAATCACTTCTCATCAGATTTGAAAAAAGACACTGAAAA 1140
Db 1743 ACACCTCCTTTCTAAATTGCAATCACTTCTCATCAGATTTGAAAAAAGACACTGAAAA 1801
QY 1141 CCAATTACTTGAATAAAATTTGACAGTAGTGTGTTTACCATCCGCTCAATTACCCAAAGAC 1200
Db 1802 CCAATTACTTGAATAAAATTTGACAGTAGTGTGTTTACCATCCGCTCAATTACCCAAAGAC 1861
QY 1201 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 1260
Db 1862 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 1921
QY 1261 TCACATAGATTTCTCCTCCGTGAGTAACTGAGTAACTGATATACAAATTTATGCTTCCAAAGTA 1320
Db 1922 TCACATAGATTTCTCCTCCGTGAGTAACTGAGTAACTGATATACAAATTTATGCTTCCAAAGTA 1981
QY 1321 TGAACACATCCCAATCAGAAAAAGGTTATCAT 1351
Db 1982 TGAACACATCCCAATCAGAAAAAGGTTATCAT 2012

RESULT 6
AAC66903
ID AAC66903 standard; cDNA; 2641 BP.
XX AC AAC66903;
XX DT 27-MAR-2001 (first entry)
XX DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
XX KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
XX inflammation; reproductive disorder; cardiovascular disorder;
XX immune disorder; musculoskeletal disorder; developmental disorder;
XX gastrointestinal disorder; cell proliferation disorder; ss.
OS Homo sapiens.
XX PN WO200068380-A2.
XX PD 16-NOV-2000.
XX PF 10-MAY-2000; 2000WO-US12811.
XX PR 11-MAY-1999; 99US-0133643.
XX PR 23-AUG-1999; 99US-0150409.

PA (INCY-) INCYTE GENOMICS INC.
XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
PI Azimzai Y;
XX
XX WPI: 2001-007395/01.
DR P-PSDB; AAB27236.
XX
XX Isolated polynucleotide encoding extracellular matrix or
PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT preventing disorders associated with expression of EXMAD such as
PT proliferative, immune and genetic disorders -
XX
PS Claim 4; Page 121-122; 129pp; English.
XX
XX The present invention provides the protein and coding sequences for 25
CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC useful in the prevention and treatment of cancers, cell proliferation,
CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC gastrointestinal disorders and inflammation.
XX
SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;
Query Match 98.7%; Score 1337.4; DB 22; Length 2641;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATGCACACAGATTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 60
DB 912 ATGCAACAAGATTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 971
QY 61 ATGGAAGAATGAAGCTAGTACTAGACTCCAAGCATGGTCAGTCATCAAGAAATTTCACAATA 120
DB 972 ATGGAAGAATGAAGCTAGTACTCCAAGCATGGTCAGTCATCAAGAAATTTCACAATA 1031
QY 121 CTACAAGGTCCACCGGCCGCCAGGGTCCCAAGAGGTGACAGAGGATCCAGGAGACCCCTC 180
DB 1032 CTACAAGGTCCACCGGCCGCCAGGGTCCCAAGAGGTGACAGAGGATCCAGGAGACCCCTC 1091
QY 181 GGGCCCAACTGGCNACAAAGGACAGAAAGGAGAGAGGGGAGGCGCTGGACCTGGCCCT 240
DB 1092 GGGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGGGGAGGCGCTGGACCTGGCCCT 1151
QY 241 GCGGGTGAGAGAGGGCCCAATTGGACAGCTGTGCTCCCGGAGAGCGTGGCGCAAGGA 300
DB 1152 GCGGGTGAGAGAGGGCCCAATTGGACAGCTGTGCTCCCGGAGAGCGTGGCGCAAGGA 1211
QY 301 TCTAAAGGCTCCACGGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 360
DB 1212 TCTAAAGGCTCCACGGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 1271
QY 361 GGGCCCACTGGGAGACCCAGGCCGCCCGGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
DB 1272 GGGCCCACTGGGAGACCCAGGCCGCCCGGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAG 1331
QY 421 CAGGGCCCTCCTGGCTTCAGGGACTTCAGGGCACCCTGGGGAGGCTGGGGTGCCTGGA 480
DB 1332 CAGGGCCCTCCTGGCTTCAGGGACTTCAGGGCACCCTGGGGAGGCTGGGGTGCCTGGA 1391
QY 481 CTTGCGGGAGCTGCAGGCTTGGCTGGGTTACGAGGATGCCAGGCCCCCAAGGCCGCCGCC 540
DB 1392 CTTGCGGGAGCTGCAGGCTTGGCTGGGTTACGAGGATGCCAGGCCCCCAAGGCCGCCGCC 1451
QY 541 GGGCCCTCCTGGCCCAATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCACCCG 600
DB 1452 GGGCCCTCCTGGCCCAATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCACCCG 1511
QY 601 GCACCGGAGACAAATGGCTGGCCGCTCACTGGAGAACTTCACAGACAAATGCTACTAT 660

DB 1512 GCACCGGAGACAATAGCTGCCCGCCTCACTGGAGAACTTCACAGACAAATGCTACTAT 1571
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGAGCTTTTCTGTGGAAGACAGTCTTCA 720
DB 1572 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGAGCTTTTCTGTGGAAGACAGTCTTCA 1631
QY 721 CATCTTGTGTTTTCATAAACAAGTACAGAGAACAGCAATGGATATAAACAAGATGGTAGGG 780
DB 1632 CATCTTGTGTTTTCATAAACAAGTACAGAGAACAGCAATGGATATAAACAAGATGGTAGGG 1691
QY 781 AGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAGATGAATGGAAGTGGCTG 840
DB 1692 AGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAGATGAATGGAAGTGGCTG 1751
QY 841 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGATAACTGGGTCAT 900
DB 1752 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGATAACTGGGTCAT 1811
QY 901 GGCATGGGCCAGGAGAGACTGTGCTGGGTGATTTATGCTGGGCAGTGGAAACGATTTTC 960
DB 1812 GGCATGGGCCAGGAGAGACTGTGCTGGGTGATTTATGCTGGGCAGTGGAAACGATTTTC 1871
QY 961 CAATGTGAAGAGCTCAATAACTTCAATTCGAAAAAGACAGGAGACAGTACTGTCATCT 1020
DB 1872 CAATGTGAAGAGCTCAATAACTTCAATTCGAAAAAGACAGGAGACAGTACTGTCATCT 1931
QY 1021 GCATATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGTCTCAAGGCAAGG 1080
DB 1932 GCATATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGTCTCAAGGCAAGG 1991
QY 1081 ACACCTCTTTCTAATTTGCATCACCCTTCTCATCAGATTGAAAAAAGAAAGCACTCAAAA 1140
DB 1992 ACACCTCTTTCTAATTTGCATCACCCTTCTCATCAGATTG-AAAAAAGAAAGCACTCAAAA 2050
QY 1141 CCAATTTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCAAGAC 1200
DB 2051 CCAATTTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCAAGAC 2110
QY 1201 TTGGGAACATAAATGTTCCCGAGGGTGATGCTGATTTTCAATTTGTCACATGGACTGAA 1260
DB 2111 TTGGGAACATAAATGTTCCCGAGGGTGATGCTGATTTTCAATTTGTCACATGGACTGAA 2170
QY 1261 TCACATAGATTCTCTCCCGTCCGTCAGTAACCGTGGGATTTATACAATTTATGTTCCCAAGTA 1320
DB 2171 TCACATAGATTCTCTCTCCCGTCCGTCAGTAACCGTGGGATTTATACAATTTATGTTCCCAAGTA 2230
QY 1321 TGAACACATCCCAATCAGAAAAAGGTTATCAT 1351
DB 2231 TGAACACATCCCAATCAGAAAAAGGTTATCAT 2261
RESULT 7
AAV55746
ID AAV55746 standard; cDNA; 2930 BP.
XX
AC AAV55746;
XX
DT 23-MAR-1999 (first entry)
XX
DE Human secreted protein clone bv227_1 coding sequence.
XX
KW Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 67..693
FT /*tag= a
XX

PN W09855614-A2.
XX 10-DEC-1998.
XX 01-JUN-1998; 98WO-US11210.
XX 29-MAY-1998; 98US-0087255.
PR 04-JUN-1997; 97US-0868696.
PR 04-JUN-1997; 97US-0868697.
PR 04-JUN-1997; 97US-0868698.
PR 04-JUN-1997; 97US-0868698.
PR 04-JUN-1997; 97US-0868899.
PR 04-JUN-1997; 97US-0868900.
PR 04-JUN-1997; 97US-0869191.
PR 04-JUN-1997; 97US-0869192.
PR 04-JUN-1997; 97US-0869193.
PR 04-JUN-1997; 97US-0869194.
XX (GEMY) GENETICS INST INC.
PA Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
PI McCoy JM, Racie LA, Spaulding V, Treacy M;
XX WPI: 1999-059912/05.
DR P-P5DB; AAW73628.
XX New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT retina, adult placenta or adult uterus cDNA libraries
XX Claim 26; Page 87-88; 127pp: English.
XX This sequence encodes a human secreted protein of the invention.
CC This DNA sequence was isolated from a human adult brain cDNA
CC library, and was designated clone bv227.1. The DNAs and proteins
CC are predicted to have biological activities which would make them
CC suitable for treating, preventing or ameliorating medical conditions in
CC humans and animals, although no supporting data is given. Suggested
CC activities include nutritional sources or supplements, immune
CC stimulating or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The DNAs
CC are also stated to be useful for gene therapy. A host cell transfected
CC with the DNA, or its subfragments and variants is useful for recombinant
CC production of the human secreted protein clones.
XX
SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;
Query Match 98.6%; Score 1335.8; DB 20; Length 2930;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1348; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 ATGCACACAGATTTGATGAGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB 1266 ATGCCAACAGATTTGATGAGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 1325
QY 61 ATGGAAGAAATCAAGCTAGTACCTCCAGCATGTTGAGTCTCATCAAGAATTTTACAATA 120
DB 1326 ATGGAAGAAATCAAGCTAGTACCTCCAGCATGTTGAGTCTCATCAAGAATTTTACAATA 1385
QY 121 CTACAAGTTCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGATCCACAGGACCCCT 180
DB 1386 CTACAAGTTCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGATCCACAGGACCCCT 1445
QY 181 GCGCCAACTGGCAACAGGACAGAAAGAGAGAGGGGAGCCTGGACCACTGGCCCT 240
DB 1446 GCGCCAACTGGCAACAGGACAGAAAGAGAGAGGGGAGCCTGGACCACTGGCCCT 1505
QY 241 GCGGGTGAGAGAGGCCCAATTTGACACAGCTGGTCCCGGAGAGCGTGGCGGCAAGGA 300
DB 1506 GCGGGTGAGAGAGGCCCAATTTGACACAGCTGGTCCCGGAGAGCGTGGCGGCAAGGA 1565

QY 301 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 360
DB 1566 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 1625
QY 361 GCGCCCAAGTGGGAGCCAGGCCCGCCGGCCACACAGGAAAGAGGACTCCCGGCCCT 420
DB 1626 GCGCCCAAGTGGGAGCCAGGCCCGCCGGCCACACAGGAAAGAGGACTCCCGGCCCT 1685
QY 421 CAGGGCCCTCCTGGCTTCCAGGGACTTCCAGGCACCGTTGGGAGCCTGGGTGGCTTGA 480
DB 1686 CAGGGCCCTCCTGGCTTCCAGGGACTTCCAGGCACCGTTGGGAGCCTGGGTGGCTTGA 1745
QY 481 CCTCGGGACTGCCAGGCTTGGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCCC 540
DB 1746 CCTCGGGACTGCCAGGCTTGGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCCC 1805
QY 541 GGCCTCTCGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCACCCCG 600
DB 1806 GGCCTCTCGGCCCATCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCCACCCCG 1865
QY 601 GCACCGGAGGACAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660
DB 1866 GCACCGGAGGACAATGAGTGGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1925
QY 661 TTTTCAGTTGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCA 720
DB 1926 TTTTCAGTTGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCA 1985
QY 721 CATCTGTTTTCATAAACACTAGAGAGAACAGCAATGGATAAATAACAGATGTTAGG 780
DB 1986 CATCTGTTTTCATAAACACTAGAGAGAACAGCAATGGATAAATAACAGATGTTAGG 2045
QY 781 ACAGAGAGCACTGGATCGGCTCCAGACTCAGAGCGTGAAGTGAAGTGGAGTGGCTG 840
DB 2046 ACAGAGAGCACTGGATCGGCTCCAGACTCAGAGCGTGAAGTGAAGTGGAGTGGCTG 2105
QY 841 GATGGACATCTCCAGACTTACAAAATTTGGAAGCTGGACAGCCGGTAACCTGGGTCAT 900
DB 2106 GATGGACATCTCCAGACTTACAAAATTTGGAAGCTGGACAGCCGGTAACCTGGGTCAT 2165
QY 901 GGCATGGCCAGGAGAGAACTGTGCTGGTTGATTTATGCTGGGAGTGGACGATTTTC 960
DB 2166 GGCATGGCCAGGAGAGAACTGTGCTGGTTGATTTATGCTGGGAGTGGACGATTTTC 2225
QY 961 CAATGTGAAGAGCTCAATAACTTTCATTCGAAAGACAGGAGAGACAGTACTGTCTCT 1020
DB 2226 CAATGTGAAGAGCTCAATAACTTTCATTCGAAAGACAGGAGAGACAGTACTGTCTCT 2285
QY 1021 GCATTATAACGAGCTGTGATGGATCACAATGAGCAAAATTTTCAGCTCTCAAGGCAAGG 1080
DB 2286 GCATTATAACGAGCTGTGATGGATCACAATGAGCAAAATTTTCAGCTCTCAAGGCAAGG 2345
QY 1081 ACACCTCTTCTTAATTCATCACCCTTCTCATCAGATTGAAAAAAGCACTCAAAA 1140
DB 2346 ACACCTCTTCTTAATTCATCACCCTTCTCATCAGATTGAAAAAAGCACTCAAAA 2404
QY 1141 CCAATTTACTGAAAAAATTTGACAGTACTGTTTTTACCATCCGTCATTTACCAAGAC 1200
DB 2405 CCAATTTACTGAAAAAATTTGACAGTACTGTTTTTACCATCCGTCATTTACCAAGAC 2464
QY 1201 TTGGGAACATAAATGTTCCCGAGGGTGATGCTGATTTTTCATTTGTCACATGGACTGAA 1260
DB 2465 TTGGGAACATAAATGTTCCCGAGGGTGATGCTGATTTTTCATTTGTCACATGGACTGAA 2524
QY 1261 TCACATAGATTCTCCTCCGTCAGTAACCGTGGATTAACAATTTATGCTTCCAAAGTA 1320
DB 2525 TCACATAGATTCTCCTCCGTCAGTAACCGTGGATTAACAATTTATGCTTCCAAAGTA 2584
QY 1321 TGAACACCTCCCAATCAGAAAAAGGTTTATCAT 1351
DB 2585 TGAACACCTCCCAATCAGAAAAAGGTTTATCAT 2615

RESULT 8
AAI60628
ID AAI60628 standard; cDNA; 2181 BP.
XX
AC AAI60628;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4617.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM41472.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4617; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;

Query Match 98.1%; Score 1328.8; DB 22; Length 2181;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 1 ATGCAACAGATTGATGAGGTGCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 60

Db 662 ATGCAACAGATTGATGAGGTGCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 721
QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAGCAGCTGGTCAGCTCATCAAGAAATTTACAATA 120
Db 722 ATGGAAGAAATGAAGCTAGTAGACTCCAGCAGCTGGTCAGCTCATCAAGAAATTTACAATA 781
QY 121 CTAAAGGTTCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 180
Db 782 CTAAAGGTTCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 841
QY 181 GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAAAGGGGAGCTGGACCACCTGGCCCT 240
Db 842 GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAAAGGGGAGCTGGACCACCTGGCCCT 901
QY 241 GCGGGTGAGAGAGGCGCAATTTGGACAGCTGTGCCCGGAGAGCGTGGCGCAAGGA 300
Db 902 GCGGGTGAGAGAGGCGCAATTTGGACAGCTGTGCCCGGAGAGCGTGGCGCAAGGA 961
QY 301 TCTAAAGGCTCCACGGGCCCCAAAGGCTCCGCTGGTTCCTCGGAAAGCCGCCCTCAG 360
Db 962 TCTAAAGGCTCCACGGGCCCCAAAGGCTCCGCTGGTTCCTCGGAAAGCCGCCCTCAG 1021
QY 361 GGCCCAAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGGCCCT 420
Db 1022 GGCCCAAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGGCCCT 1081
QY 421 CAGGGCCCTCCTGGCTTCAGGAGCTTCAGGCGCACCGTTGGGGAGCGCTGGGCTGGA 480
Db 1082 CAGGGCCCTCCTGGCTTCAGGAGCTTCAGGCGCACCGTTGGGGAGCGCTGGGCTGGA 1141
QY 481 CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAAGGATGCCAGGCCCCCAAGGCCCCCCC 540
Db 1142 CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAAGGATGCCAGGCCCCCAAGGCCCCCCC 1201
QY 541 GGCCCTCCTGGCCCATCAGAGGCGGTGGTCCCTGGCCCTGCAGAAATGAGCCCAAGCCCG 600
Db 1202 GGCCCTCCTGGCCCATCAGAGGCGGTGGTCCCTGGCCCTGCAGAAATGAGCCCAAGCCCG 1261
QY 601 GCACCGGAGGACAAATGGCTGCCCGCTCCTGGAAGAACTTACACAGCAATGCTACTAT 660
Db 1262 GCACCGGAGGACAAATGGCTGCCCGCTCCTGGAAGAACTTACACAGCAATGCTACTAT 1321
QY 661 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGAGCTTTTGTGGAAGCAAGCTTCA 720
Db 1322 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGAGCTTTTGTGGAAGCAAGCTTCA 1381
QY 721 CATCTGTTTTCATAAACAAGAGGAAACCAATGGATAAAAAACAGATGGTAGGG 780
Db 1382 CATCTGTTTTCATAAACAAGAGGAAACCAATGGATAAAAAACAGATGGTAGGG 1441
QY 781 AGAGAGAGCCACTGGATCGCCCTCAGAGACTCAGAGCGTGAAATGAATGAAGTGGCTG 840
Db 1442 AGAGAGAGCCACTGGATCGCCCTCAGAGACTCAGAGCGTGAAATGAATGAAGTGGCTG 1501
QY 841 GATGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTCAT 900
Db 1502 GATGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTCAT 1561
QY 901 GGCCATGGCCAGGAGAAAGCTGTGCTGGGTTGATTATGCTGGGAGTGGAAAGATTTTC 960
Db 1562 GGCCATGGCCAGGAGAAAGCTGTGCTGGGTTGATTATGCTGGGAGTGGAAAGATTTTC 1621
QY 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGAGAGGAGACAGTACTGTCTATCT 1020
Db 1622 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGAGAGGAGACAGTACTGTCTATCT 1681
QY 1021 GCATTATAACGAGCTGTGATGGGATCAGATGAGCAAAATTTTTCAGCTCTCAAGGCAAGG 1080
Db 1682 GCATTATAACGAGCTGTGATGGGATCAGATGAGCAAAATTTTTCAGCTCTCAAGGCAAGG 1741
QY 1081 ACACCTCTTTCTTAATTTGATCAGCTTCTCATCAGATTTGAAAAAAGAGGAGCTCAAAA 1140


```
QY 541 GGCCTCTCGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAATGAGCAACCCCG 600
Db 1806 GGCCTCTCTCGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAATGAGCAACCCCG 1865
QY 601 GCACCGGAGGACAATGGCTGCGCCCTCACTGGAGAGACTTCACAGACAAATGCTACTAT 560
Db 1866 GCACCGGAGGACAATGGCTGCGCCCTCACTGGAGAGACTTCACAGACAAATGCTACTAT 1925
QY 661 TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAAGACAAGTCTTCA 720
Db 1926 TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAAGACAAGTCTTCA 1985
QY 721 CATCTGTTTTTCATAAACAATAGAGAGGAACAGCAATGATGATAAAAAACAGATGGTAGGG 780
Db 1986 CATCTGTTTTTCATAAACAATAGAGAGGAACAGCAATGATGATAAAAAACAGATGGTAGGG 2045
QY 781 AGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGACGCTGAAATGAATGGAAGTGGCTG 840
Db 2046 AGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGACGCTGAAATGAATGGAAGTGGCTG 2105
QY 841 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCCGATAACTGGGTCTAT 900
Db 2106 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCCGATAACTGGGTCTAT 2165
QY 901 GGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTC 960
Db 2166 GGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTC 2225
QY 961 CAATGTGAAGAGCTCAATAACTTCATTTCCGAAAAGACAGGAGACAGTACTGTCATCT 1020
Db 2226 CAATGTGAAGAGCTCAATAACTTCATTTCCGAAAAGACAGGAGACAGTACTGTCATCT 2285
QY 1021 GCATTATAACGGAGCTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAAGGCAAGG 1080
Db 2286 GCATTATAACGGAGCTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAAGGCAAGG 2345
QY 1081 ACATCTCTTTCTAATGTCAATACCTTCTCATCAGATTGAAAAAAGCACTGAAAA 1140
Db 2346 ACATCTCTTTCTAATGTCAATACCTTCTCATCAGATTGAAAAAAGCACTGAAAA 2404
QY 1141 CCAATTACTGAAAAAAATTTGACAGTAGTGTGTTTACCATCCGTCATTACCCAAAGAC 1200
Db 2405 CCAATTACTGAAAAAAATTTGACAGTAGTGTGTTTACCATCCGTCATTACCCAAAGAC 2464
QY 1201 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 1260
Db 2465 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 2524
QY 1261 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTAACAATTTATGTCCTTCCAAAGTA 1320
Db 2525 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTAACAATTTATGTCCTTCC-AAAGTA 2583
QY 1321 TGAACACTCCCAATCAGAAAAAGGTTATCAT 1351
Db 2584 TGAACACTCCCAATCAGAAAAAGGTTATCAT 2614
```

```
RESULT 10
ID ABA97932
XX ABA97932 standard; cDNA; 2262 BP.
AC ABA97932;
XX ABA97932;
XX ABA97932;
DT 25-APR-2002 (first entry)
XX Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
DE Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
KW rheumatoid arthritis; gene; ss.
XX Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
```

```
FT CDS 33..2262
FT /*tag= a
FT /product= "scavenger receptor-like protein"
XX
PN JP2001340089-A.
PD 11-DEC-2001.
XX
PF 08-DEC-2000; 2000JP-0375066.
XX
PR 27-MAR-2000; 2000JP-0090772.
XX
PA (SHIO ) SHIONOGI & CO LTD.
XX
DR WPI; 2002-144965/19.
DR P-PSDB; ABB08642.
XX
New scavenger receptor-like protein for diagnosis, prevention and
treatment of autoimmune disease, such as rheumatoid arthritis -
Claim 3; Fig 1; 38pp; Japanese.
CC The invention relates to a human scavenger receptor-like protein. The
protein is useful as a target molecule for diagnosis, prevention and
treatment of autoimmune diseases such as rheumatoid arthritis.
XX
SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
Query Match 75.9%; Score 1029; DB 24; Length 2262;
Best Local Similarity 100.0%; Pred. No. 1.8e-236;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAACAAGATTTCATGAGTTCGAGGTTAGACACTGAAAGTAGCCAACTTATCAGTGATT 60
Db 1234 ATGCAACAAGATTTCATGAGTTCGAGGTTAGACACTGAAAGTAGCCAACTTATCAGTGATT 1293
QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACAATA 120
Db 1294 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACAATA 1353
QY 121 CTACAGAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACGAGGACCCCT 180
Db 1354 CTACAGAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACGAGGACCCCT 1413
QY 181 GGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGAGCCCTGGACACCTGGCCCT 240
Db 1414 GGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGAGCCCTGGACACCTGGCCCT 1473
QY 241 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCGCCGAGAGAGGCTGGCGCAAGGA 300
Db 1474 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCGCCGAGAGAGGCTGGCGCAAGGA 1533
QY 301 TCTAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 360
Db 1534 TCTAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 1593
QY 361 GGCCCCAGTGGGAGCCAGGCCCGCCCGGGCCCAACAGGAGAGGAGGACTCCCGGCCCT 420
Db 1594 GGCCCCAGTGGGAGCCAGGCCCGCCCGGGCCCAACAGGAGAGGAGGACTCCCGGCCCT 1653
QY 421 CAGGGCCCTCTCTGGCTTCAGGGACTTCAGGGACACGCTTGGGGAGGCTGGGCTGGA 480
Db 1654 CAGGGCCCTCTCTGGCTTCAGGGACTTCAGGGACACGCTTGGGGAGGCTGGGCTGGA 1713
QY 481 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAGGCTGCCAGGCCCAAGGCCCGCC 540
Db 1714 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAGGCTGCCAGGCCCAAGGCCCGCC 1773
QY 541 GGCCCTCTCTGGCCCAATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAATGAGCAACCCCG 600
Db 1774 GGCCCTCTCTGGCCCAATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAATGAGCAACCCCG 1833
QY 601 GCACCGGAGGACAATGGGCTGCCCGCTCACTTGAAGAACTTTCACAGACAAATGCTACTAT 660
```

Db 1834 GCACGGAGGACAAATGGCTGCCGCTCAGTGGAGAACTTTCACAGACAAAATGCTACTAT 1893
Oy 661 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGGAAGACAAGTCTTCA 720
Db 1894 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGGAAGACAAGTCTTCA 1953
Oy 721 CATCTTGTGTTTATTAACACTAGAGAGGAGCAAGCAATGGATAAAAAACAGATGGTAGGG 780
Db 1954 CATCTTGTGTTTATTAACACTAGAGAGGAGCAAGCAATGGATAAAAAACAGATGGTAGGG 2013
Oy 781 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGCTCAGAGCTGAAATGAATGAAGTGGCTG 840
Db 2014 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGCTCAGAGCTGAAATGAATGAAGTGGCTG 2073
Oy 841 GATGGGACATCTCCAGACTACAAAAATGGAAGCTGGACAGCCGATAACTGGGGTTCAT 900
Db 2074 GATGGGACATCTCCAGACTACAAAAATGGAAGCTGGACAGCCGATAACTGGGGTTCAT 2133
Oy 901 GGCCATGGGCCAGGAGAAAGACTGTGCTGGTGTGATTTATGCTGGGCACTGGAACGATTC 960
Db 2134 GGCCATGGGCCAGGAGAAAGACTGTGCTGGTGTGATTTATGCTGGGCACTGGAACGATTC 2193
Oy 961 CAATGTGAAGACGTCATTAACCTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCT 1020
Db 2194 CAATGTGAAGACGTCATTAACCTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCT 2253
Oy 1021 GCATTATAA 1029
Db 2254 GCATTATAA 2262

RESULT 11
AAH43037
ID AAH43037 standard; cDNA; 2637 BP.
XX AC AAH43037;
XX DT 15-OCT-2001 (first entry)
XX DE Nucleotide sequence of a human scavenger receptor.
XX KW Human; scavenger receptor; SRCL-Pl; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 92..2320
FT FT /*tag= a
FT FT /product= "scavenger receptor"
XX PN WO200159107-A1.
XX PD 16-AUG-2001.
XX PF 08-FEB-2001; 2001WO-JP00874.
XX PR 14-FEB-2000; 2000JP-0035155.
XX PR 10-OCT-2000; 2000JP-0309068.
XX PA (FUSO) FUSO PHARM IND LTD.
XX PI Wakamiya N;
XX DR WPI; 2001-497076/54.
XX DR P-PSDB; AAG63347.
XX PT New Scavenger receptor proteins SRCL-Pl with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation -
XX

PS Claim 6; Page 88-93; 118pp; Japanese.
XX The present sequence encodes a human scavenger receptor, designated
CC SRCL-Pl. The SRCL-Pl polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SO Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;

Query Match 66.1%; Score 895.2; DB 22; Length 2637;
Best Local Similarity 81.7%; Pred. No. 9.5e-222;
Matches 1100; Conservative 0; Mismatches 228; Indels 18; Gaps 5;

Oy 1 ATGCAACAAGATTTGATGAGGTTCGAGTTAGACACTGAAGTAGCCAACTTATCATGTGATT 60
Db 1292 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCATGTGTT 1351
Oy 61 ATGGAAGAATAAGTAGTAGACTCCAAAGCATGCTCAGCTCATCAAGAATTTTACAATA 120
Db 1352 ATGGAAGAATAAGTAGTAGACTCCAAAGCATGCTCAGCTCATCAAGAATTTTACCATT 1411
Oy 121 CTACAAGTCTCCACGGGCCCGAGGGTCCAAAGAGTGCACAGAGTCCAGGGAGCCCT 180
Db 1412 CTACAAGTCTCTGCGCCCGACAGAGTCCAAAGGTGCACAGAGTCTCAGGGACCCT 1471
Oy 181 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCTTGACACCTGGCCCT 240
Db 1472 GGTCCAACCTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGAGCTTGGTCCACCTGGCCCT 1531
Oy 241 GCGGGTGAGAGAGGCCCAATTTGGACCAAGTGGTGTCCCGCGGAGAGCGTGGCGGCAAGA 300
Db 1532 GCGGGTGAGAGAGGCCCAATTTGGACCAAGTGGTGTCCCGCGGAGAGCGTGGCGGCAAGA 1591
Oy 301 TCTAAAGGCTCCGAGGGGCCCAAGGCTCCGCTGGTGTCCCTGGGAAGCCCGGCCCTCAG 360
Db 1592 TCCAAAGGCTCACAGGTTCCCAAGGATCTCGTGGTCCCGAGGAGCGTGGCCCTCA 1651
Oy 361 GGCCCACTGGGGACCCAGGCCCCCGGCCCAACAGGAGAGAGGAGTCTCCCCGGCCCT 420
Db 1652 GGACCTAGTGGGGACCCAGGACCACAGTGTCCACAGGCAAGGATGACTTCCCTGGCCCT 1711
Oy 421 CAGGCCCCCTCTGGCTTCCAGGGACTTCAGGGACCTTGGGGACCTTGGGGTGGCTTGA 480
Db 1712 CAGGCCCCCTCTGGCTTCCAGGGACTTCAGGGACCTTGGGGTGGAGTACCTTGA 1771
Oy 481 CTGCGGGACTGCCAGGCTTGCCTGGGTACCGAGCATGCCAGGCCCCCAAGGGCCCCCCC 540
Db 1772 CTGCGGGGTTGCCAGGCTTGCAGGGGTGCCAGGATGCTTGGGCTTAAAGGACCACT 1831
Oy 541 GGCCCTCTGCGCCCATCAGAGCGGTGTGCCCCCTGGCCCTGSCAGATGAGCCAAACCCG 600
Db 1832 GGCCCTTCCAGGCCCTCAGGAGCAATGGAGCATTGGCTTGCAGAAATGAACCAACCCCA 1891
Oy 601 GCACGGAGGACATGGCTGCCGCTCAGTGGAGAGACTTCACAGACAAATGCTACTAT 560
Db 1892 GCATCAGAGGTCAACGGATGTCCGCTCAGTGGAGAACTTTCACAGATAAATGCTACTAT 1951
Oy 661 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGGAAGACAAGTCTTCA 720
Db 1952 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCTAAGCTTTTCTGTGGAAGACAAATCTTCC 2011
Oy 721 CATCTTGTGTTTATTAACACTAGAGAGGAGCAAGCAATGGATAAAAAACAGATGGTAGGG 780
Db 2012 CATCTGTTTATTAACACTCAAGAGAGCAAGCAATGGATAAAAAACAGATGGTAGGG 2071
Oy 781 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGCTGGAAGTGAATGAAGTGGCTG 840
Db 2072 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGAGGAGGAGGAAATGGAAGTGGCTA 2131

Db 772 CTGTGTTTCATTAACACATAGAGAGGACAGCAATGGATAAAAAACACAGATGTTAGGGAGA 831

Qy 784 GAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTGGAT 843

Db 832 GAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTGGAT 891

Qy 844 GGGACATCTCAGACTCAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGC 903

Db 892 GGGACATCTCAGACTCAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGC 951

Qy 904 CATGGCCAGGAGAGACTGTCTGGGTTGATTTATGCTGGGCACTGGAACGATTTCCAA 963

Db 952 CATGGCCAGGAGAGACTGTCTGGGTTGATTTATGCTGGGCACTGGAACGATTTCCAA 1011

Qy 964 TGTGAAGCGTCAATAACTTCATTTGCGAAAAAGACAGGGAGACAGT 1010

Db 1012 TGTGAAGCGTCAATAACTTCATTTGCGAAAAAGACAGGGAGACAGT 1058

RESULT 13

AAH43054

ID AAH43054 standard; DNA; 2256 BP.

XX AC AAH43054;

XX DT 15-OCT-2001 (first entry)

XX DE Nucleotide sequence of a human scavenger receptor.

XX KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;

XX KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;

XX KW low density lipoprotein; ss.

XX OS Homo sapiens.

XX FH Key

FT CDS

FT Location/Qualifiers

FT 74..1936

FT /*tag= a

FT /product= "scavenger receptor"

PN WO200159107-A1.

XX PD 16-AUG-2001.

XX PF 08-FEB-2001; 2001WO-JP00874.

XX PR 14-FEB-2000; 2000JP-0035155.

XX PR 10-OCT-2000; 2000JP-0309068.

XX PA (FUSO) FUSO PHARM IND LTD.

XX PI Wakamiya N;

XX DR WPI; 2001-497076/54.

XX DR P-PSDB; AAG63350.

XX PT New Scavenger receptor proteins SRCL-P1 with collectin-like structure,

XX PT useful for treatment and diagnosis of diseases associated with oxidized

XX PT low-density lipoprotein accumulation -

XX PS Claim 4; Page 105-109; 118pp; Japanese.

XX CC The present sequence encodes a human scavenger receptor, designated

XX CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They

XX CC are useful in clarifying the functions of macrophages and basal

XX CC immunity. They are also useful in the treatment, prevention, and basal

XX CC and investigation of diseases such as arteriosclerosis, diabetic

XX CC complications, bacterial infection and restenosis following angioplasty,

XX CC which are associated with accumulation of oxidized low density

XX CC lipoprotein and the binding of advanced glycation end-products into

XX CC cells.

XX SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Query Match 54.9%; Score 743.8; DB 22; Length 2256;
Best Local Similarity 91.9%; Pred. No. 1.6e-182;
Matches 797; Conservative 0; Mismatches 67; Indels 3; Gaps - 1;

Qy 489 ACTGCCAGGCTTGGCTGGGTACACAGGATGCCAGGCCCCAAGGGCCCCCGGCCCTCC 548

Db 1393 ACTAAGGTCCACCGGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCAGGAGCCCC 1452

Qy 549 TGGCCCATCAGAGGGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGCACCGGA 608

Db 1453 TGGCCCAACTGCAACAAGGGACAGAAAGGAGAGGGGGAGCGCTGGAC---CACCTGG 1509

Qy 609 GGACAATGGCTGCCCGCTCACTGGAAGAACTTTCACAGACAAATCTCTACTATTTTCAGT 668

Db 1510 CCTCGGGCTGCCCGCTCACTGGAAGAACTTTCACAGACAAATCTCTACTATTTTCAGT 1569

Qy 669 TGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGCAAGTCTTCACATCTTGT 728

Db 1570 TGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGCAAGTCTTCACATCTTGT 1629

Qy 729 TTTCAATAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAG 788

Db 1630 TTTCAATAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAG 1689

Qy 789 CCACTGGATCGGCTTCACAGACTCAGAGCGTGAAATGAATGGAATGGCTGGATGGGAC 848

Db 1690 CCACTGGATCGGCTTCACAGACTCAGAGCGTGAAATGAATGGAATGGCTGGATGGGAC 1749

Qy 849 ATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGG 908

Db 1750 ATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGG 1809

Qy 909 GCCAGGAGAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGGAAACGATTTCCAATGTGA 968

Db 1810 GCCAGGAGAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGGAAACGATTTCCAATGTGA 1869

Qy 969 AGACGTCAATAACTTCATTTGCGAAAAAGACAGAGGAGACAGTACTGTCTGCAATTATA 1028

Db 1870 AGACGTCAATAACTTCATTTGCGAAAAAGACAGAGGAGACAGTACTGTCTGCAATTATA 1929

Qy 1029 ACGGACTGTGATGGATCACATGAGCAAAATTTTCAGTCTCAAAGGCAAGAGCACTCCT 1088

Db 1930 ACGGACTGTGATGGATCACATGAGCAAAATTTTCAGTCTCAAAGGCAAGAGCACTCCT 1989

Qy 1089 TTCTAATTGCATCACCTTCTCATCAGATTGAAAAAAGCACTGAAAAACCAATTAC 1148

Db 1990 TTCTAATTGCATCACCTTCTCATCAGATTGAAAAAAGCACTGAAAAACCAATTAC 2049

Qy 1149 TCAAAAAAATTGACAGCTAGTGTGTTTTTACCATCCGTCAATACCAAGACTTGGGAAC 1208

Db 2050 TCAAAAAAATTGACAGCTAGTGTGTTTTTACCATCCGTCAATACCAAGACTTGGGAAC 2109

Qy 1209 TAAATGTTCCCGGGTGATATGCTGATTTTCATTTGTCACATGGATGATACATAG 1268

Db 2110 TAAATGTTCCCGGGTGATATGCTGATTTTCATTTGTCACATGGATGATACATAG 2169

Qy 1269 ATTCTCTCCCTCAGTACCGGTGGATTATACAAATTTATGTTCCAAAGTATGAAACAC 1328

Db 2170 ATTCTCTCCCTCAGTAAACCGGTGGATTATACAAATTTATGTTCCAAAGTATGAAACAC 2229

Qy 1329 TCCAATCAGAAAAAGGTTATCATCCG 1355

Db 2230 TCCAATCAGAAAAAGGTTATCATCCG 2256

RESULT 14

AA571131/c

ID AA571131 standard; cDNA; 493 BP.

XX AC AA571131;

XX DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #6935.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI
XX
XX
XX P-PSDB; ABC06944.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 6935; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 493 BP; 121 A; 116 C; 99 G; 157 T; 0 other;
SQ

Query Match
Best Local Similarity 30.5%; Score 413; DB 23; Length 493;
Matches 482; Conservative 0; Mismatches 5; Indels 6; Gaps 6;

QY 812 CAGACGGTGAATAATGAATGAAGTGGCTGGATGGGACATC-TCCAGACTACA-AAATTTG 869
|||||
DB 493 CAGACGGTGAATAATGAATGAAGTGGCTGGATGGGACATC-TCCAGACTACA-AAATTTG 434
|||||
QY 870 GAAAGCTGGAC-AGCGGGATAACT-GGGTCATGCCATGGGCCAGGAGACACTGTGCT 927
|||||
DB 433 GAAAGCTGGCAAGCCGGATGAATGAATGGGCTCATGCCATGGGCCAGGAGACACTGTGCT 374
|||||
QY 928 GGGTTGATTATGCTGGCAGTGGAAACCATTTTCCAATGTGAAGACGTCATTAACCTTCATT 987
|||||
DB 373 GGGTTGATTATGCTGGCAGTGGAAACCATTTTCCAATGTGAAGACGTCATTAACCTTCATT 314
|||||
QY 988 TCGGAAAAGACAGGAGACACTACTGTCATTCATTAACGACACTGTGATGGGATCA 1047
|||||

Db 313 TCGGAAAAGACAGGAGACAGTACTGTCTATCTGCATTATATAACGAGCTGTGTATGGGATCA 254
QY 1048 CATGAGCAAAATTTTCAGCTCTCAAAGGCAAGGACACTCTTTCTTAATTGCATCACCTTC 1107
|||||
Db 253 CATGAGCAAAATTTTCAGCTCTCAAAGGCAAGGACACTCTTTCTTAATTGCATCACCTTC 194
QY 1108 TCATCAGATTGAAAAAAGACACTGAAACCAATTAATGAAAAAATTTGACAGCT 1167
|||||
Db 193 TCATCAGATTGAAAAAAGACACTGAAACCAATTAATGAAAAAATTTGACAGCT 134
QY 1168 AGTGTGTTTTTACCATCCGTCATTACCCAAAAGACTTTGGGAACATAAATGTTCCCA 1226
|||||
Db 133 AGTGTGTTTTTACCATCCGTCATTACCCAAAAGACTTTGGGAACATAAATGTTCCCA 74
QY 1227 GATATGCT-GATTTTCATTGTGCACATGGACATGAATFACATAGATTCTCCTCCGTCAGTA 1285
|||||
Db 73 GATATGCTGGATTTCATTGTGCACAGGACTGAATFACATAGATTCTCCTCCGTCAGTA 14
QY 1286 ACCGTGCGATTAT 1298
|||||
Db 13 CCGGTGCGATTAT 1
RESULT 15
ABAL17153
ID ABAL17153 standard; DNA; 5041 BP.
XX
AC ABAL17153;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 9484.
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparasitic; antitickling; antianaemic; antirheumatic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiparasitic; antidiabetic; antitickling; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS Homo sapiens.
XX
XX WO200159063-A2.
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225260.
PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0242221.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-541565/60.
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -
PS Disclosure; SEQ ID NO 9484; 1701pp + Sequence Listing; English.
XX The invention relates to novel genes (AB11004-AB21534) and proteins
XX (AB14678-AB18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 5041 BP; 1480 A; 1074 C; 1103 G; 1384 T; 0 other;

```
Query Match      24.5%; Score 332; DB 22; Length 5041;
Best Local Similarity 99.7%; Pred. No. 1.3e-75;
Matches 343; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1008 AGTACTGTCTATCTGCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTC 1067
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1739 AGTACTGTCTATCTGCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTC 1798

QY 1068 TCAAAGGCAAAAGGACACTCCTTTCTAATTGCATCAGCTTCTCATCAGATTGAAAAA 1127
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1799 TCAAAGGCAAAAGGACACTCCTTTCTAATTGCATCAGCTTCTCATCAGATTG-AAAAA 1857

QY 1128 AAAGCACTGAAACCAATTAAGCAAAATTAAGCAAAATTAAGCAAAATTAAGCAAA 1187
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1858 AAAGCACTGAAACCAATTAAGCAAAATTAAGCAAAATTAAGCAAAATTAAGCAAA 1917

QY 1188 ATTACCCAAAGACTTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTGTG 1247
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1918 ATTACCCAAAGACTTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTGTG 1977

QY 1248 CACATGGACTGAATCACATAGATTCTCTCCGTCAGTAACCGTGGATTATACAAATTAT 1307
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1978 CACATGGACTGAATCACATAGATTCTCTCCGTCAGTAACCGTGGATTATACAAATTAT 2037

QY 1308 GTCTTCCAAAGTATGGAACACTCCCAATCAGAAAAAGGTTATCAT 1351
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2038 GTCTTCCAAAGTATGGAACACTCCCAATCAGAAAAAGGTTATCAT 2081
```

Search completed: March 21, 2003, 02:19:58
Job time : 300.408 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 06:54:47 ; Search time 3616.46 Seconds
(without alignments)
4401.885 Million cell updates/sec

Title: US-09-763-712A-2

Perfect score: 2969

Sequence: 1 MYSHNVIMNLNLTLQVQ.....EDVNNFICEKDRETVLSSAL 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/usptc_pool/US09763712/runat_14032003_100949_18101/app_query.fasta_1.1877
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_WAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn_1_1.6828_runat_14032003_100949_18101 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2369	100.0	2983	9 AB005145	AB005145 Homo sapi
2	2363	99.8	2641	6 AX047353	AX047353 Sequence
3	2363	99.8	3058	9 AB038518	AB038518 Homo sapi
4	2765	93.1	2637	10 AB078434	AB078434 Mus muscu
5	2751	92.7	3291	10 AB038519	AB038519 Mus muscu
6	2693	90.7	4330	9 AB052103	AB052103 Homo sapi
7	2688	90.5	2005	6 AX454442	AX454442 Sequence
8	2688	90.5	2005	6 AX490920	AX490920 Sequence
9	2309	77.8	1886	9 HSM802985	AL713657 Homo sapi
10	1239	41.7	169088	2 AC016128	AC016128 Homo sapi
11	1239	41.7	178022	2 AP001022	AP001022 Homo sapi
12	1239	41.7	187635	2 AP001022	AP001022 Homo sapi
13	1239	41.7	188439	2 AP000935	AP000935 Homo sapi
14	1212	40.8	188255	2 AP000939	AP000939 Homo sapi
15	1168	39.3	182029	2 AC114677	AC114677 Mus muscu
16	1168	39.3	193208	2 AC102618	AC102618 Mus muscu
17	1021.5	34.4	130763	2 AC112416	AC112416 Rattus no
18	991	33.4	71044	2 AC024368	AC024368 Homo sapi
19	981	33.0	182029	2 AC114677	AC114677 Mus muscu
20	981	33.0	193208	2 AC102618	AC102618 Mus muscu
21	842	28.4	188255	2 AP000939	AP000939 Homo sapi
22	765	25.8	3636	9 AB007829	AB007829 Homo sapi
23	765	25.8	3685	6 E32511	E32511 Scavenger r
24	765	25.8	3810	6 E32509	E32509 Scavenger r
25	671	22.6	2215	10 BC026446	BC026446 Mus muscu
26	621	20.9	188637	2 AL845306	AL845306 Danio rer
27	491.5	16.6	1265	10 RATSPD	M81231 Rat pulmona
28	483.5	16.3	3508	3 AF053538	AF053538 Alvinella
29	481.5	16.2	1183	10 BC003705	BC003705 Mus muscu
30	481.5	16.2	1253	10 MUSSPD	L40156 Mus musculu
31	474.5	16.0	1410	6 AX334792	AX334792 Sequence
32	474.5	16.0	1410	9 HSMRNAPD	X65018 H.sapiens m
33	474	16.0	1454	4 BTLSPD	X75911 B.taurus mR
34	473.5	15.9	1284	4 AF509590	AF509590 Bos tauru
35	473.5	15.9	1301	9 BC022318	BC022318 Homo sapi
36	471	15.9	2377	9 BC008760	BC008760 Homo sapi
37	471	15.9	5676	6 E07265	E07265 cDNA encodi
38	471	15.9	5676	6 HUMCALV	D90279 Human mRNa
39	471	15.9	7138	9 HUMPAIV	M76729 Human pro-a
40	465	15.7	6114	10 CRUPAIV	M76730 Chinese ham
41	462.5	15.6	4428	6 AX146422	AX146422 Sequence
42	462.5	15.6	4428	6 AX146424	AX146424 Sequence
43	462.5	15.6	4990	10 CRUCOLVIA	L06863 Cricetus
44	462	15.6	2777	5 GU07973	U07973 Gallus gall
45	462	15.6	5575	5 AF137273	AF137273 Gallus ga

ALIGNMENTS

... RESULT 1

AB005145
LOCUS 2983 bp mRNA linear PRI 21-NOV-2001
DEFINITION Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1 GI:17026100
KEYWORDS
SOURCE
ORGANISM
Homo sapiens female tissue_lib:placenta cDNA to mRNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuo,H., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.
TITLE The membrane-type collectin CL-P1 is a scavenger receptor on
vascular endothelial cells
J. Biol. Chem. 276 (47), 44222-44228 (2001)
JOURNAL
MEDLINE 21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani,K.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail:ohtanik@asahikawa-med.ac.jp,
Tel:+81-166-68-2393, Fax:+81-166-68-2399)
FEATURES
source
1..2983
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
1..2983
/gene="CL-P1"
71..2299
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDVFABEEVQSFYKRFQIEGTQCTKCKNNWALKFSILLY
ILCALLTTVAIGYKVKMDNVTGGMETSRQYVDKLTAVESDLKLGQDTGKAI
STNSLSTFRSDILDROOLREITEKTSKNDTLEKQASGDALVDROSQLETKENN
SELIITVAKTQVANGYVNLQDDSVLQGNLQNMYSHNVVIMNMLNLTQVOQRN
LITNLRSDVTSQAIRIKNDQFNQLOKQKIDTDLKQVQSLQTLAANNLSALA
KANNDTIDMNSISFTGMENITTSANEQNLKDLQDLKHAENRTAIFKNQLEE
RFQLETDIVNIISYTAHLRLTLTSLNNEVRTCTDTLTKHDDTLTSLNNTLANI
RLDSVLRMOODLMRSRLDEVANLSVIMEKLVDSKHGOLIKNFTILOGPPGPRGP
RGDRSGQPPGPTGNKGOKGEGEPGPGAGEGPIGAPGPGERGGKSGSGPK
LPGVPMGPKPMPGPPGPGSGAVPLALQNEPTPAPEDNGCPPHWKNFTDKCYFSVE
KEIFEDAKLFCEDKSSHLVFINTREEQWIKQMVGRSHWIGLTDSERENWKWDG
TSPDYKNNKAGQPDNMGHGPEDCAGLIYAGQWDFQCEVDVNNFICEKDRQTVLSS
AL"

BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN
Alignment Scores:
Pred. No.: Length: 2983
Score: 2969.00 Matches: 547
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-763-712a-2 (1-547) x AB005145 (1-2983)
QY 1 MetTyrSerHisAsnValVallMetAsnLeuAsnAsnLeuThrGlnValGln 20
DB 656 ATGTATTCATCAATGTGTCATCATGAACTCAACACCTGAACCTGACCCAGGTGCAG 715
QY 21 GlnArgAsnLeuThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
DB 716 CAGAGGAACCTCATACGAACTCTGAGCGGTCTGTGGATGACACAGCCAGGCTATCCAG 775

QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
DB 776 CCAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGGACACG 835
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
DB 836 GATTGGCTCAAGGAGAAAGTCAGAGCTTTCAGACAGCTGGCTGCCAACAACTCTCGGTG 895
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
DB 896 GCCAAGCCCAACACGACACCTCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 955
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
DB 956 CAGATGGAGAACATCACCACCTATCTCTCAAGCCCAAGCAGACCACTGAAAGACCTGCAG 1015
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
DB 1016 GACTTACACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCACCACTGGAGGAACGC 1075
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
DB 1076 TTCCAGCTCTTTGAGACGGATATTTGGAACATCATTTAGCAATATCATGTTACAGCCCCAC 1135
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
DB 1136 CACCTGGCGGACCTGACCAAGCAATCTAAATGAAGTCAGGACCACCTTGCACAGATACCTT 1195
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
DB 1196 ACCAAACACACAGATGATCTGACCTCTCTTGAATAATACCTTGCCCAACATCCGTTTGGAT 1255
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
DB 1256 TCTGTTTCTCTCAGGATGCACACAGATTGTAGAGGTCAGGTTAGACACTGAAGTAGCC 1315
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
DB 1316 AACTTATCATGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGTCAGCTCATC 1375
QY 241 LysAsnPheThrIleLeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGly 260
DB 1376 AAGAATTTTACAATACTACAAGGTCACCGGGCCCGGGGTCCTCAAGAGGTGACAGAGA 1435
QY 261 SerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGluPro 280
DB 1436 TCCAGGGGACCCCTGGCCCACTGGCACACAGGGGACAGAGGAGAGAGGGGAGCCT 1495
QY 281 GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu 300
DB 1496 GGACCACTGGCCCTCGGGTGGAGAGGCCCCAATTTGGACCAGCTGGTCCCCCGGAGAG 1555
QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
DB 1556 CGTGGCGCAAGGATCTAAAGGCTCCCGAGGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGG 1615
QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGlu 340
DB 1616 AAGCCCGGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCACAGGCAAGAG 1675
QY 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
DB 1676 GGACTCCCGCGCCCTCAGGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCCTGGGGAG 1735
QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
DB 1736 CCTGGGTGCTCTGGACCTCGGGACTTGCACAGCTTGCCTGGGGTACCAGGCAATGCCAGGC 1795
QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
DB 1796 CCCAAGGGCCCCCGGGCCCTCTCTGGCCCTCAGGAGCGGTGGTGGCCCTGGCCCTGCAG 1855

```

Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpPlyAsnPheThr 420
Db 1856 AATGAGCCAAACCCGGCACCAGGACAAATGGCTGCCCGCCTCACTGGAAGAACTTCACA 1915
Qy 421 AspLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1916 GACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGT 1975
Qy 441 GluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db 1976 GAAGACAAGTCTCACATCTGTTTTCATAAACACTAGAGGAGACACAAATGGATAAAA 2035
Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2036 AAACAGATGCTAGGAGAGAGACCACTGGATCGGCCCTCACAGACTCAGAGCGTGAANA 2095
Qy 481 GluTrpLysTyrLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyClnPro 500
Db 2096 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGAGCGCG 2155
Qy 501 AspAsnTrpGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2156 GATAACTGGGTGCATGGCCATGGCCAGGAGAGACTGTGCTGGTTCATTTATGCTGG 2215
Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2216 CAGTGGAAACGATTTCCAAATGTGAAGACGTCAATAACTTCAATTTGCGAAAAAGACAGGG 2275
Qy 541 ThrValLeuSerSerAlaLeu 547
Db 2276 ACAGTACTGTATCTGCATTA 2296

RESULT 2
AX047353
LOCUS AX047353 2641 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2641)
AUTHORS Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R.,
Lu,D.A. and Azimzai,Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
INCYTE Genomics, Inc. (US)
FEATURES
source
1..2641
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CBI"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN

Alignment Scores:
Pred No.: 2,64e-95 Length: 2641
Score: 2963.00 Matches: 546
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.80% Indels: 0
DB: 6 Gaps: 0

US-09-763-712A-2 (1-547) x AX047353 (1-2641)
Qy 1 MetTrpSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 297 ATGTATTTCTCATATGTGGTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGGAG 356
Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40

```

```

Db 357 CAGAGGAACCTCATCATCAGGAATCTGCAGCGGTGTGTGGATGACACAAAGCAGGCTATTCAG 416
Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 417 CGAATCAGAAGACGACTTTCAAAATCTGCAGCAGGTTTTTCTCAAGCCCAAGAGGACAGC 476
Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
Db 477 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACCTCTGCGTTG 536
Qy 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 537 GCCAAAGCCACAACAGCACACCTTGAGGATATGAAACCCAGCTCAACTCATTCACAGT 596
Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 597 CAGATGGAGACATCACCACATCTCTCAAGCCACAGGAGAGAACTGAAAGACCTGGCAG 656
Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 657 GACTTACACAAAGATGCAGAGAAATAGAACGACATCAAGTTCAACCAACTGGAGGAACGC 716
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrThrAlaHis 160
Db 717 TTCCAGCTCTTTGAGACCGGATATTGTGAACATCATTAGCAATATCAGTTACACAGCCCAC 776
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValAlaArgThrThrCysThrAspThrLeu 180
Db 777 CACCTGCGGACGCTGACCAGCAATCTAATGAAGTCAGGACCCTTGCACAGATACCTT 836
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 837 ACCAAACACACAGATGATCTGACCTCTTGAATAATAATCCCTGGCCCAACATCCGTTGGAT 896
Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 897 TCTGTTCTCTCAGGATGCAACAAAGATTTGATGAGTTCGAGTTAGAGTTCAGACTCAG 956
Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 957 AACTTATCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCACAGCATGGTCAGCTATC 1016
Qy 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
Db 1017 AAGAAATTTTACAATACTACAAGTCCACCGGCCCCAGGGTCCCAAGAGTGCACAGAGCA 1076
Qy 261 SerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
Db 1077 TCCCAGGAGCCCTCTGGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGCGCT 1136
Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
Db 1137 GGACCACCTGGCCCTGCGGGTGAGAGAGGCCCAATTTGACAGCTGGTCCCCCGGAGAG 1196
Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1197 CGTGGCGGCAAAAGGATCTAAAGGCTCCAGGGGCCCAAAAGGCTCCCGTGGTCCCTGGG 1256
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProGlyLysGlu 340
Db 1257 AAGCCCGCCCTCAGGGGCCCAAGTGGGACCCAGGCCCGCCCGGGCCACAGAGCAAGAG 1316
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1317 GGACTCCCGCGCCCTCAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGCTGGGGAG 1376
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1377 CCTGGGGTGCCTGGACCTCGGGGACTGCAGGCTTGCCTGGGGTACCAGGCATGCCAGC 1436
Qy 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGln 400
Db 1437 CCCAAGGGCCCCCGGCCCTCTCGCCCATCAGGAGCGGTGTGCTGCCCTGCCCTGCAG 1496

```

QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
|||||
Db 1497 AATGAGCAACCCGCGCAGGACAAATAGTCTGCCCGCTCACTGGGAAGAACTTCACA 1556
QY 421 AspLysCysTyTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1557 GACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTCT 1616
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db 1617 GAAGACAAGCTCTTCATCTCTTTTCATAACACTAGAGAGAAACAGCAATGGATAAAA 1676
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 1677 AAACAGATGCTAGGAGAGAGAGCCACTGATCGCGCTCACAGACTCACAGCGTGAATAAT 1736
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnPro 500
Db 1737 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCG 1796
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyraGly 520
Db 1797 GATACTGGGCTCATGCCATGGCCAGGAGAGACTGTCTGGGTGATTTATGCTGGG 1856
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 1857 CAGTGGACGATTTCAATGTGAAGCGTCAATACTTCTTGGCAAAAAGACAGGGAG 1916
QY 541 ThrValLeuSerSerAlaLeu 547
Db 1917 ACAGTACTGTCTATCTGCATTA 1937
RESULT 3
LOCUS AB038518 3058 bp mRNA linear PRI 08-MAR-2001
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type 1, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE Homo sapiens tissue_lib:Placenta cdna to mRNA.
ORGANISM
REFERENCE
AUTHORS Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
REFERENCE 2 (bases 1 to 3058)
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@bich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
source
1..3058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Placenta"
gene 1..3058
/gene="SRCL"
CDS 209..2437
/codon_start=1
/product="scavenger receptor with C-type lectin type I"
/protein_id="BAB39147.1"
/db_xref="GI:13365515"

/translation="MKDDFAEEVEVQSEYKRFQIQEQTCTCKTKNNNALKPSIILLY
ILCALLITVAILGYKVVERKMDNVTGKMETSRQTYDDKLITAVESDKLQDQTKKAI
STNSLSLTFRSDILDRLQOLREITEKSKNDKTLKLAQSDALVDROSQKLTLENN
SFLITVNTKLTQAVNGYVTNLQDTSVLQGNLONOMYSHNVVIMNLNLNLTQVOORN
LITNLQSRVDDTSQAIOIKNDQNLQFLQAKKDTMLKEKVQSLQTLAANNSALA
KANNDLEDMSQLNSFTQGMENITTSQANEONLKDLDLHKDAENRTAKFNQLEE
RFQFETDIVNIISNISTAHRLTLTSLNEVKTCTDTLTUKHDDTLTSLNNTLANI
RLDSVSLRMOODLMSRLDTEVANLSVMEEMKLVDSKHGOLIKNFTLQGGPPGRGP
RGDSGSGPPGPTNGKGKGKGPPEGPAGEPGPGPGPGPGPGPGPGPGPGPGPGPK
GSRSPGKPGQPGSDPGPGPGKGLPGQPGPGPGPGPGPGPGPGPGPGPGPGPGPG
LPGVGMGPKPGKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG
KEIFEDAKLFCEDKSSHLVFINTRBEQWIKQWVGRESHWIGLTDLSRENEWKLDG
TSPDYKKNKAGCPDNWGHGHPGEDCAGLIYAGQWDFQCEDVNNFICEKORETVLSS
AL"
polya_signal 3020..3025
/gene="SRCL"
BASE COUNT 911 a 744 c 764 g 639 t
ORIGIN
Alignment Scores:
Pred. No.: 2,97e-95 Length: 3058
Score: 2963.00 Matches: 546
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.80% Indels: 0
DB: 9 Gaps: 0
US-09-763-712A-2 (1-547) x AB038518 (1-3058)
QY 1 MetTySerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 794 ATGTATTCTCATATGTGGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAG 853
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
Db 854 CAGAGGAACCTCATCAGCAATCTGCAGCGCTCTGTGGATGACACAGCCAGGCTATCCAG 913
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
Db 914 CGAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACAG 973
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 974 GATTGGCTGAAGAGAAAGTGCAGAGCTTGCAGACACTGGCTGCCAACACTCTGCGTTG 1033
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 1034 GCCAAGCCCAACAGCACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGT 1093
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 1094 CAGATGGAGAACATCACCCTATCTCTCAAGCCCAAGCAGACAGCAACCTGAAGACCTGCAG 1153
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 1154 GACTTACAAAAGATGCGAGAGATAGAACAGCCATCAAGTTCAACCACTGGAGGAACCC 1213
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
Db 1214 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTACACAGCCCAC 1273
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1274 CACCTGGCGAGCTGACCCAGCAATCTAAATGAAGTCAGGACCACCTTGACACAGATACCTT 1333
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1334 ACCAAACACACAGATGATCTGACCTCTCTTGAATAATACCTGGCCAAACATCGCTTGGAT 1393
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1394 TCTGTTTCTCTCAGGATGCAACAAGATTTGATGAGGTGCGAGGTTAGACACTGAAGTAGCC 1453

[illegible]

Qy		41	ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr	60
Db		797	CGAATTAAAGATGATTTCCAAAATCTGCAGCAGGGTTTTTCTTCACGCCAAGAAGCACACC	856
Qy		61	AspTrpLeuLysGluLysValcIcnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu	80

```

Db      857  GATTGGCTAAAGAAAGTACAGAGCTTGCAGACATTGGCTGCCAACAACTCTGCCCTG 916
QY      81  AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db      917  GCCAAAGCCAAACATGACACCCTAGAGGATATGAATAGCAGCTCAGCTCATTCAGGT 976
QY      101  GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db      977  CAGATGGACAACATTACCACTATCTCACAGGCCAACGACAGAGCCTGAAAGACCTTCAG 1036
QY      121  AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db      1037  GACTTACACAGGATACAGAAATAGAACAGCTGTCAAGTTGAGCCAACTTGAGGAACGC 1096
QY      141  PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerIleThrAlaHis 160
Db      1097  TTCCAGGTCTTGAGACACATATTGTGAACATCATTTAGCAACATCAGCTACAGCCCAT 1156
QY      161  HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db      1157  CACCTGAGGACACGTGACCAAGCAATCTGAATGATGTTAGGACCACATGCACAGACCTTG 1216
QY      181  ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db      1217  ACCAGACACAGGATGACCTGACCTCCTTGAATAACACACTAGTCAACATCCGCTTGGAT 1276
QY      201  SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db      1277  TCTATTCTCTCAGGATGACAGACATGATGAGTCAAGTGTAGACACTGAGTGGCC 1336
QY      221  AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db      1337  AACTTATCAGTGTATGAAGAGATGAACCTGTTGACTCCAAGCACGCTCAGCTCATC 1396
QY      241  LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
Db      1397  AAGAACTTTACCACTTACAAAGTCTCTGCGCCCGAGAGTCCAAAGGTGACAGAGA 1456
QY      261  SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlyGluPro 280
Db      1457  TCTCAGGACCACCTGCTCCACTGCAACAGGGGACAGAGGAGAGAGAGCT 1516
QY      281  GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
Db      1517  GTTCCACCTGGCCCTCGCGGTGAGAGGGGCACAAATTGGACCACTCGCTCGAGAG 1576
QY      301  ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db      1577  CGTGGCAGCAAAAGGATCCAAAGGCTCACAGGTTCCCAAGGATCTCGTGGGTCCCCAGG 1636
QY      321  LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
Db      1637  AAGCTTGGCCCTCAAGACCTAGTGGGACCCAGGACCAACCACTGCTCCAGCAAGGAT 1696
QY      341  GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db      1697  GGACTCCCTGGCCCTCAGGGCCCTCTCGCTTCCAGGGACTACAGGGCACTGTGGGTGAG 1756
QY      361  ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db      1757  CTTGGAGTACCTGGACCTCGGGGGTTCAGAGCTTTCAGAGGGGTGCCAGGCATGCTGGG 1816
QY      381  ProLysGlyProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGln 400
Db      1817  CCTAAGGGACCACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCATGGCTCTGCGAG 1876
QY      401  AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db      1877  AATGAACCAACCCAGCATCAGAGGTCAACGGATGTCGCCCTCAGTGGGAAGAACTTACA 1936
QY      421  AspLysCysTyrThrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440

```

```

Db      1937  GATAAATGCTACTATATTTTCATTGGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGT 1996
QY      441  GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db      1997  GAAGACAAATCTTCCATCTCGTTTTCATAAACTCAAGAGAAGAACAGCAATGGATAAAA 2056
QY      461  LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db      2057  AAGCATACCGTGGGAGAGAAAGCCATTGGATCGGCTCACAGACTCAGAACAGGAAGC 2116
QY      481  GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
Db      2117  GAATGAAGTGGCTAGACGGGTCACTGTGATTACAAAACTGGAAAGCTGACAAACA 2176
QY      501  AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db      2177  GATACTGGGCGAGTGGCCATGGCCAGGAGAGACTGTGCTGGCTTGATTTACGCGAG 2236
QY      521  GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGlyLysAspArgGlu 540
Db      2237  CAGTGAATGACTTCCAGTGTGATGAATCAATACTTCACTTGTGAGAAGGAAGGAG 2296
QY      541  ThrValLeuSerSerAlaLeu 547
Db      2297  GCAGTACCATCATCCATATTA 2317

```

RESULT 5
LOCUS AB038519 3291 bp mRNA linear ROD 18-JAN-2002
DEFINITION Mus musculus src1 mRNA for scavenger receptor with C-type lectin,
complete cds.
ACCESSION AB038519
VERSION AB038519.1 GI:18146951
KEYWORDS
SOURCE Mus musculus cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T.
TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin
(SRC1)(1), a novel member of the scavenger receptor family
JOURNAL Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
MEDLINE 21575692
REFERENCE 2 (bases 1 to 3291)
AUTHORS Nakamura, K. and Nakamura, T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
source 1..3291
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..3291
/gene="src1"
CDS 77..2305
/gene="src1"
/codon_start=1
/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDDFAEEEVQSFYKRFGLHEGTCQTKCINNNAKFSIVLLY
ILCALLTITVAILGKYVVEKMDNVDGNETSHQYDNKLTAVESDLKLLGDOAGKKAL
STNSLSLTFRSDILRLQLOEITKTSKNDLLEKLAQNGSLVDVQSLQSLKETLQNN
SFLTITVNTKTLQAYNYTNLQDDTNVLOGNLQSQMSQSVVIMNLNLTFQVQRN
LISNLQSQVDDTSLAIQRKNDQFLQAKKDTDLKQVSLQTLAAANSALA
KANNDTLDMNSLSSTFGQMDNITTSQANEQSLKLDQLDKHDTENRTAVKFSLEE
RFQVFETDIVNIISNIVTAHLRPTLSNLDVMTCTDTLTRHDDTLTSLNPTLVNI
RLDSLSLRMOODMRKSLDTEVANLVSVMVEEMKLVDSKHGLIKNFTILQGPGRGP
KDRGSQGPPTGKNKGKGEKGPPEPPGAGEKGTIGVPDPGERGSKSGSGPK

01234567891011121314151617181920212223242526272829303132333435363738394041424344454647484950515253545556575859606162636465666768697071727374757677787980818283848586878889909192939495969798991001011021031041051061071081091101111121131141151161171181191201211221231241251261271281291301311321331341351361371381391401411421431441451461471481491501511521531541551561571581591601611621631641651661671681691701711721731741751761771781791801811821831841851861871881891901911921931941951961971981992002012022032042052062072082092102112122132142152162172182192202212222232242252262272282292302312322332342352362372382392402412422432442452462472482492502512522532542552562572582592602612622632642652662672682692702712722732742752762772782792802812822832842852862872882892902912922932942952962972982993003013023033043053063073083093103113123133143153163173183193203213223233243253263273283293303313323333343353363373383393403413423433443453463473483493503513523533543553563573583593603613623633643653663673683693703713723733743753763773783793803813823833843853863873883893903913923933943953963973983994004014024034044054064074084094104114124134144154164174184194204214224234244254264274284294304314324334344354364374384394404414424434444454464474484494504514524534544554564574584594604614624634644654664674684694704714724734744754764774784794804814824834844854864874884894904914924934944954964974984995005015025035045055065075085095105115125135145155165175185195205215225235245255265275285295305315325335345355365375385395405415425435445455465475485495505515525535545555565575585595605615625635645655665675685695705715725735745755765775785795805815825835845855865875885895905915925935945955965975985996006016026036046056066076086096106116126136146156166176186196206216226236246256266276286296306316326336346356366376386396406416426436446456466476486496506516526536546556566576586596606616626636646656666676686696706716726736746756766776786796806816826836846856866876886896906916926936946956966976986997007017027037047057067077087097107117127137147157167177187197207217227237247257267277287297307317327337347357367377387397407417427437447457467477487497507517527537547557567577587597607617627637647657667677687697707717727737747757767777787797807817827837847857867877887897907917927937947957967977987998008018028038048058068078088098108118128138148158168178188198208218228238248258268278288298308318328338348358368378388398408418428438448458468478488498508518528538548558568578588598608618628638648658668678688698708718728738748758768778788798808818828838848858868878888898908918928938948958968978988999009019029039049059069079089099109119129139149159169179189199209219229239249259269279289299309319329339349359369379389399409419429439449459469479489499509519529539549559569579589599609619629639649659669679689699709719729739749759769779789799809819829839849859869879889899909919929939949959969979989991000100110021003100410051006100710081009101010111012101310141015101610171018101910201021102210231024102510261027102810291030103110321033103410351036103710381039104010411042104310441045104610471048104910501051105210531054105510561057105810591060106110621063106410651066106710681069107010711072107310741075107610771078107910801081108210831084108510861087108810891090109110921093109410951096109710981099110011011102110311041105110611071108110911101111111211131114111511161117111811191120112111221123112411251126112711281129113011311132113311341135113611371138113911401141114211431144114511461147114811491150115111521153115411551156115711581159116011611162116311641165116611671168116911701171117211731174117511761177117811791180118111821183118411851186118711881189119011911192119311941195119611971198119912001201120212031204120512061207120812091210121112121213121412151216121712181219122012211222122312241225122612271228122912301231123212331234123512361237123812391240124112421243124412451246124712481249125012511252125312541255125612571258125912601261126212631264126512661267126812691270127112721273127412751276127712781279128012811282128312841285128612871288128912901291129212931294129512961297129812991300

TITLE	Molecular cloning and functional characterization of a human scavenger receptor with C-type lectin (SRCL), a novel member of a scavenger receptor family
JOURNAL	Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE	21092718
REFERENCE	2 (bases 1 to 4330)
AUTHORS	Nakamura,K. and Nakamura,T.
TITLE	Direct Submission
JOURNAL	Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate School of Medicine, Division of Biochemistry, Biomedical Research Center, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:knakamura@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783, Fax:81-6-6879-3789)

FEATURES	source	Location/Qualifiers
	1. .4330	/organism="Homo sapiens"
		/db_xref="taxon:9606"
gene	1. .4330	
		/gene="SRCL"
CDS	3. .1871	
		/gene="SRCL"
		/codon_start=1
		/product="scavenger receptor with C-type lectin type II"
		/protein_id="BAB39148.1"
		/db_xref="GI:13365553"
		/translation="MKDDFAEESEVPSCFKRGIHEGTQCPKCHNNWALKFSILILY ILCALLTIYVAILGYKVEKMDNVTGCMSTRQTYDDKLTAVESDALKGDDQTGKRAI STNSELSTFRSDILDRQURELEIKETSRKNDTLEKLAQSGDALVDROSQLKRELN SFLITNTQTLQAYNGYVQDQTLVLOCNLQNYSHNVIMNNLNLTQVQRN LITNQRVDDTSQAIQRIKNDQNLQVFLQAKDITDKERQVSLQTLAANSLA KANNDLEDMNSLFTGQNMENITTSQANEQLKDLQDLHKDAENRTAIKFNQLEE RFOLFETDINIIISNTSYTAHHURTLTSLNEVYRTCTDPTLTKHTDDLTSLNNTLANI RDLDSLRMOODLMRSRLDTEVANSVIMEEMKLVDSKHCOLIKNTILQGPQGP RCDRSGQPPPTGNGKQGEKEGPEPPGAGERGPIGAGPEGRGGKSGKSGQGP GSRGSEKPEQPGQSGGEGELPGQPGPEGLQGTGVEGPGVGPGRGLPG LPQVPMGPKGPPGPGPSGAVVPLALQNEPTPAPEDNSKSPSLQPGQGGSACA"
BASE COUNT	1186 a	1065 c 1035 g 1044 t
ORIGIN		
Alignment Scores:		
Pred. No.:	9, 25e-86	Length: 4330
Score:	2693.00	Matches: 547
Percent Similarity:	50.70%	Conservative: 0
Best Local Similarity:	50.70%	Mismatches: 0
Query Match:	90.70%	Indels: 532
DB:	9	Gaps: 1

US-09-763-712A-2 (1-547) x AB052103 (1-4330)

Qy	1	MctTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln	20
Db	588	ATGTATTCTCATAATGTGGTGCATCATGAACCTCAACAACCTGAACCTGACCCAGGTCGAC	647
Qy	21	GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln	40
Db	648	CAGAGGAACCTCATCAGAACTTCGACGGCTCTGTGGATGCACAACGACCGCATCCAG	707
Qy	41	ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuclnAlaLysLysAspThr	60
Db	708	CGAATCAAGAACGACTTTCAAAATCTGCACAGGTTTTCTTCAAGCCCAAGAGGACACG	767
Qy	61	AspTrpLeuLysGlnLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu	80
Db	768	GATTGGCTGAAGGAGAAGTGCAAAGCTTGCACACACTGGCTGCCAACCACTCTGGCGTG	827
Qy	81	AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly	100
Db	828	GCCAAAGCCCAACACACACCCCTGGAGGATATCAACAGCCAGCTCAACTCATTCACAGT	887
Qy	101	GlnMetGlnAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln	120
Db	888	CAGATGGAGAATCATCACACTATCTCTCAAGCCCAACGACGACAACTGAAAGACCTGCAG	947

Qy	121	AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg	140
Db	948	GACTTACACAAAGATGCAGAGAATAAGACAGCATCAAGTTCAACCAACTGGAGAACGC	1007
Qy	141	PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis	160
Db	1008	TTCAGCTCTTTTCAGACGGATATTGTGAACATCATTTAGCAATATCAGTTACACAGCCAC	1067
Qy	161	HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu	180
Db	1068	CACCTCGGAGCGTGCACAGCAATCTAAATGAAGTCAGACCACCTTGCACAGATACCCCT	1127
Qy	181	ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp	200
Db	1128	ACCAACACACACATGATCTGACCTCTTGAATAATACCTTGGCCAACTCCGTTTGAT	1187
Qy	201	SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla	220
Db	1188	TCTGTTTCTCTCAGGATGCAACAAGATTGTATGAGTGCAGGTATAGACACTCAAGTAGCC	1247
Qy	221	AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle	240
Db	1248	AACCTATCAGTGATATGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATC	1307
Qy	241	LysAsnPheThrIleLeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGly	260
Db	1308	AAGATTTTACATACTACAGGTCACCGGCCCCAGGGGTCCAAGAGGTGCACAGAGGA	1367
Qy	261	SerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro	280
Db	1368	TCCAGGGAGCCCTCGGCCAACTGGCAACAAGGACAGAAAGAGAGAGAAGGGGAGCCT	1427
Qy	281	GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlyC	300
Db	1428	GGACCACTCGGCCCTCGGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCCCGGAGAG	1487
Qy	301	ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly	320
Db	1488	CGTGGCGGCAAAAGATCTAAGGCTCCCAAGGCCCCCAAGGCTCCCGTGGTTCCCTGGG	1547
Qy	321	LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGlu	340
Db	1548	AAGCCCGGCCCTCAGGGCCCACTGGGGACCCAGGCCCCCGGGGCCCAACAGCAAGAG	1607
Qy	341	GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyC	360
Db	1608	GGACTCCCGGGCCCTCAGGGCCCTCTGCTCTCCAGGACTTCAGGGCACCGCTGGGGAG	1667
Qy	361	ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly	380
Db	1668	CCTGGGTGCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGC	1727
Qy	381	ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln	400
Db	1728	CCCAAGGGCCCCCGGCCCTCTCGGCCCATCAGGAGGGTGTGTGCCCTGGCCCTGCAG	1787
Qy	401	AsnGluProThrProAlaProGluAspAsn	410
Db	1788	AATGAGCCAAACCCCGGCACCGGAGGACAATAGTATGATCCAAAGCCCTCCCTCCAGCCAGG	1847
Qy	410	-----	410
Db	1848	GGACAGGCAGTGCATGCTTCAGACTCAGCCAAACCCCACTGTGTGTGGCCCCCACCATG	1907
Qy	410	-----	410
Db	1908	TTATTTTGTGTTTGTGTTTGTGTTTGTGAGACAGGGTCTCACTCTGTGTGCCACAGCTGGAGT	1967
Qy	410	-----	410
Db	1968	GCAGTGGCGTGATCTCGGTCTACTGCAACCCCAACATCCCGGGTTCAAGTGATTTCTCCTG	2027
Qy	410	-----	410

Db 2028 CTTAGCTCCAGTAGCTGGGAATAAGCGTGTGCCCCACAGCCAGCTAATTTTG 2087
Qy 410 ----- 410
Db 2088 TATTTTTAGTAGAGGGTTTCACAGTGTGGGAGGGTGGTTTCAAACTTCTGGCCTC 2147
Qy 410 ----- 410
Db 2148 AAGTGATCCACCTACCTAGTTTCCCAAAGTGTGGGAATACAGGTGTGAACCAACACCTT 2207
Qy 410 ----- 410
Db 2208 TTGTCCTCCCGGTCATATTATTTAAACAGCAAAGGCATATAGGGGTTGGCACAGGACC 2267
Qy 410 ----- 410
Db 2268 CTGCTGCCAGAACAGGCTCCATACCAGGCATGTCTCTCTTAACGCCATTTGGGTTT 2327
Qy 410 ----- 410
Db 2328 TTTAGTGTCTATTTCTTTTATGTATGTGTGTATGTATGGACAGGGTCTCACTCTGT 2387
Qy 410 ----- 410
Db 2388 CACCCAGGCTGGAATGACGAAGTGAAGTAGTGCAAAACATAGCTCATTTGACCT 2447
Qy 410 ----- 410
Db 2448 CCTGGCTTAGTGATCTCCCACTCAGACTCCTCCGTAGCTGGGACACAGGGGTATGCG 2507
Qy 410 ----- 410
Db 2508 CACCACACCTGGCTAATTTTATTTTCTTTTGTAGAACAGGGTCTTGCTATGTTGC 2567
Qy 410 ----- 410
Db 2568 CGAGCTGGTCTGAACCTCTGGCTCAAGCAACCTCTGCTCCTCAGTCTTGCAAGTGC 2627
Qy 410 ----- 410
Db 2628 TGGGATCCAGGTGGAGCCACTGCACCTGGCTCTTGACACTTGTATTTATATAGAGCCA 2687
Qy 410 ----- 410
Db 2688 GGCCTCAGAAGACAAACAAATGCAAAACATTTCTTCAGGGAAGCCTGCTGGCGCCT 2747
Qy 410 ----- 410
Db 2748 GAGCGCTCTCCCTCGTTTATCATCTCCTCAGTCTGGAGATAGATGGGCTCCCTCGGTTAGC 2807
Qy 410 ----- 410
Db 2808 TCCATCAACATGTGATGGAACCCCTCTGCTATGTAGTACTACATTTAATTTCTTTTGAG 2867
Qy 410 ----- 410
Db 2868 AAAGGTGTAGATGTTACGGTCCCTCTAGTCTTTTGATAATTAACCTTTGGCAAACT 2927
Qy 410 ----- 410
Db 2928 CGAATGTTTGTAGCAACCAAGCCTATTTTCTCTCCTTGACCAACCGCTATCTAC 2987
Qy 410 ----- 410
Db 2988 TTGATTTTGGGTGAGGCGCGGGTTATTTTGTGTATCCGTGAGCCCATTTTGGTGC 3047
Qy 410 ----- 410
Db 3048 TGGGTCTCACATCCAAGTGTCTGGGTTCCGTATGATTCAGACTCAGCGTGAACATC 3107
Qy 410 ----- 410

Db 3108 TGTGCTCCCTGCTCTTGTGTTGCTTCTGTCTCATTTCTGTCAATGTGTGAACGCGATCC 3167
Qy 410 ----- 410
Db 3168 AAACACAGAAGNACAAAGACCCAGTCCAAGATCCATCCAGATTATGTGGTTTCGGCTTGT 3227
Qy 410 ----- 410
Db 3228 GCTCTGCAGTCCACTTGTGGTTCATGCTTGTGCGAGCAGCTCCGTGAACGTGAGCTGTGG 3287
Qy 410 ----- 410
Db 3288 GCCTCGTCAGGAGCGCTCAGGCTGCTCCAGCGGACACAGTTTGGCCCTCTGAAACA 3347
Qy 410 ----- 410
Db 3348 CAGCTGAAGAAAGGACAGAGTCTTCTGTGATCTTTTAAATCAATGTTTTCACATTTTGA 3407
Qy 411 -----GlyCysProHisTrpLysAsnPherThrAspLysCysTyrTyrPheSerVal 428
Db 3408 TTCTTAGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTT 3467
Qy 429 GILYLSGLULPheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuVal 448
Db 3468 GAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGT 3527
Qy 449 PheLeAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer 468
Db 3528 TTCATAACACTAGAGAGGACAGCANTGGATATAAAACACAGATGGTAGGGAGAGAGCC 3587
Qy 469 HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr 488
Db 3588 CACTGGATCGGCTCACAGACTCAGAGCGTGAATGAAATGAATGGAAGTGGCTGGTGGACA 3647
Qy 489 SerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAspTrpGlyHisGlyHisGly 508
Db 3648 TCTCCAGACTACAAAATTTGGAAGCTGGACAGCGGATAACTGGGGTCATGGCCATGGG 3707
Qy 509 ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
Db 3708 CCAGGAGAAGACTGTGCTGGGTTGATTATGCTGGCAGTGGAGCAATTTCCAATGTGAA 3767
Qy 529 AspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 3768 GACGTCAATACTTCATTTGCAAAAGACAGGAGACAGTACTGTCTCATCTGCATTA 3824

RESULT 7

AX454442

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

AX454442 2005 bp DNA linear PAT 06-JUL-2002
Sequence 27 from Patent WO208284.
AX454442
AX454442.1 GI:21713845
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Baker, K.P., Ferrara, N., Gerber, H., Gerritsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.
Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US); Baker, Kevin P. (US); Ferrara, Napoleone
(US); Gerber, Hanspeter (US); Gerritsen, Mary E. (US); Goddard,
Audrey (US); Godowski, Paul J. (US); Gurney, Austin L. (US);
Hillan, Kenneth J. (US); Marsters, Scot A. (US); Pan, James (US);
Paoni, Nicholas F. (US); Stephan, Jean-Philippe F. (US);
Watanabe, Colin K. (US); Williams, P. Mickey (US); Wood, William
I. (US)
Location/Qualifiers

```

source      1. .2005
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 7,47e-86 Length: 2005
Score: 2688.00 Matches: 495
Percent Similarity: 99.60% Conservative: 2
Best Local Similarity: 99.20% Mismatches: 2
Query Match: 90.54% Indels: 1
DB: 6 Gaps: 0

US-09-763-712A-2 (1-547) x AX454442 (1-2005)

Qy 49 LeuGlnGlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGln 58
Db 8 GTCCGGCAGGTTTTCCTCAAGCCAGAGAGACACGGATGGCTGAAGGAGAAAGTGCAG 67
Qy 69 SerLeuGlnThrLeuAlaAlaAsnSerAlaLeuAlaLysAlaAsnAsnAspThrLeu 88
Db 68 AGCTTGCAGACACTGGCTGCCAACAACTCTGCCTGGCCAAACCCAAACGACACCCGTG 127
Qy 89 GluAspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIle 108
Db 128 GAGGATATGAACGCCAGCTCAACTCATTCACAGGTCAGATGGAGAACATCACCACTATC 187
Qy 109 SerGlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsn 128
Db 188 TCTCAAGCCAAACGAGCAGAACCTTGAAGACCTCGAGGACTTACACAAAGATCGAGAAAT 247
Qy 129 ArgThrAlaIleLysPheAsnGlnLeuGluAluArgPheGlnLeuPheGluThrAspIle 148
Db 248 AGACAGCCATCAAGTTCACCAACTGGAGAACGCTTCCAGCTCTTTGACAGCGATATT 307
Qy 149 ValAsnIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsn 168
Db 308 GTGAACATCATTTAGCAATATACGTTACACAGCCACACCTCGGGGACGCTGACCAACAAT 367
Qy 169 LeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThr 188
Db 368 CTAATGAAGTCAGGACCACTTGACAGATACCTTACCAACACACACATGATCTGCACC 426
Qy 189 SerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGln 208
Db 427 TCCTTGAATAATACCTTGGCCACATCCGTTGGATCTGTTTCTCTCAGGATGCAACA 486
Qy 209 AspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGlu 228
Db 487 GATTTGTATGAGTTCGAGGTTAGACACTGAAGTAGCACTTATCAGTGATATATGGAAGAA 546
Qy 229 MetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGly 248
Db 547 ATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAATTTTACAATACTACAAGT 606
Qy 249 ProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThr 268
Db 607 CCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCCAGGGACCCCTGGCCCACT 666
Qy 269 GlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGlu 288
Db 667 GGCAACAAGGCACAGAAAGAGAGAGAGGGGGAGCCCTGGACCCTGGCCCTGCGGGTGA 726
Qy 289 ArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGly 308
Db 727 AGAGGCCCAATTGGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGC 786
Qy 309 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 328
Db 787 TCCAGGGCCCCCAAGAGCTCCCGTGGTTCCTTCCCTTGGGAAGCCCGGCCCTCAGGGCCCCAGT 846
Qy 329 GlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyPro 348

```

[illegible]

Db	967	CTGCCAGGCCTTGCTCGGGGTACCAAGCATGCCAGGCCCAAGAGGCCCGCCCTCCT	1026
QY	389	GlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu	408
Db	1027	GGCCCATCAGAGGGTGGTGGCCCTCGCCCTGCAGAAATGAGCAACCCCGCAGGAG	1086
QY	409	AspAsnGlyCysProProHisTrpLysAsnPhetrAspLysCysTyTrpPheSerVal	428
Db	1087	GACAATAGCTGCCCGCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTT	1146
QY	429	GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuVal	448
Db	1147	GAGAAAGAAATTTTGAGGATGCAAGCTTTCTGTGAAGCAAGCTTCACATCTTGTT	1206
QY	449	PheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer	468
Db	1207	TTATAAACACTAGAGAGAACACCAATGGATAAAAAACAGATGGTAGGAGAGAGAGC	1266
QY	469	HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr	488
Db	1267	CACCTGGATCGCCTCACAGACTCAGAGCGTGAAGTGAATGAAGTGGCTGGATGGGACA	1326
QY	489	SerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly	508
Db	1327	TCCTCAGACTACAAAAATGGAAAGCTGGACAGCGGATAACTGGGTCATGGCCATGG	1386
QY	509	ProGlyGluAspCysAlaGlyLeuIleTyAlaGlyGlnTrpAsnAspPheGlnCysGlu	528
Db	1387	CCAGGAGAACACTGTCTGGGTGATTTATCTGGCAGTGGAAAGTTCCTCAATGTGAA	1446
QY	529	AspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu	547
Db	1447	GACGTCAATAACTTCATTTGCGAAAAAGACAGGAGAGACTGCTCATCTGCATTA	1503
RESULT 9			
LOCUS	HSM802985		
DEFINITION	Homo sapiens mRNA; cDNA DKFp547G1215 (from clone DKFZp547G1215).		
ACCESSION	AL713657		
VERSION	AL713657.1	GI:19584339	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1886)		
AUTHORS	Koehler,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY		
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp547G1215) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: cloneerzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/ Location/Qualifiers		
FEATURES			
source	1..1886		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="18p11.3"		
	/clone="DKFZp547G1215"		
	/tissue_type="brain"		
	/clone_lib="547 (synonym: hfbr1). Vector pSport1; host DH10B; sites NotI + SalI"		
gene	1..1886		
	/dev_stage="fetal"		
CDS	<1..1256		

```
/gene="DKF2p547G1215"
/Note="Similarity to asialoglycoprotein receptor (Mus musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKFNQLEERFQFETDIVNIISNISTYTAHLRLTLTSLNLEVRT
TCTDTLTKHTDLSLNNLTANIRLDSVSRMQODLMRSRLDTEVANLSVIMEKMLV
DSKHQLIKNTFTILQPPGPRGDRSGOPPGPTGNKGOKGKGPDPGPPGAGERG
PIGPAGPGEGRGSGKSGSPGKSGSPGPGQSSGDPGPGPKGKGLPGQGP
PGFQGLQGTGEPGPGRGLPLGPGVEGMPGKPGPPGPGPSANVPLALQNEPTPA
PEDKCPHWNFTDKCYFSEKEIFEDAKLFCDEKSSHLVFINTREEQWIKRQWV
GRESHWIGLTDSEENENKWLDTGSPDYKNKAGQPDNWMHGHPGDECAGLIYAGOW
NDFQCEDNNFTICEKRETVLSSAL"
polyA_signal
1838..1843
/gene="DKF2p547G1215"
1869
polyA_site
/gene="DKF2p547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Alignment Scores:
Pred. No.: 1,02e-72 Length: 1886
Score: 2309.00 Matches: 416
Percent Similarity: 99.76% Conservative: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 77.77% Indels: 0
DB: 9 Gaps: 0
US-09-763-712a-2 (1-547) x HSM802985 (1-1886)
QY 131 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 150
Db 3 GCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGAGCGATATTGTGAAC 62
QY 151 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrIleuThrSerAsnLeuAsn 170
Db 63 ATCATTAGCAATATCATGTTACAGACCCACCCCTCGGAGCGCTGACCAACATCTAAAT 122
QY 171 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 190
Db 123 GAAGTCAGGACCACTTGCACATACCTTACCAACACACACAGATGATCTGACCTCCTTG 182
QY 191 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 210
Db 183 AATAATACCTCGCCAAACATCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 242
QY 211 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 230
Db 243 ATGAGTCCAGGTTAGACACTGAAGTAGCACTTATCATGATTATGGAAGAAATGAAG 302
QY 231 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 250
Db 303 CTAGTAGACTCCAGCATGGTCAGCTCATCAAGAAATTTTACAACTACTCAAGGTCCACCG 362
QY 251 GlyProArgGlyProArgGlyAspArgGlySerGlnGlnGlyProGlyProThrGlyAsn 270
Db 363 GGCCCCAGGGTCCAGAGAGTGCAGAGAGATCCAGGAGACCCCTGGCCCAACTGGCAAC 422
QY 271 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 290
Db 423 AAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCACTTGGCCCTCGCGGTGAGAGAGGC 482
QY 291 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 310
Db 483 CCAATTGGACCACTGGTCCCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCACAG 542
QY 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlnGlyProSerGlyAsn 330
Db 543 GGCCCCAAGAGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGCTCCAGTGGGAC 602
QY 331 ProGlyProProGlyProProGlyLysGlyGluLeuProGlyProGlnGlnGlyProGly 350
```

```
Db 603 CGAGCCCGCCGGGGCCACCAGGCAAGAGGAGACTCCCGGGCCCTCAGGGCCCTCTGGC 662
QY 351 PheGlnGlyLeuGlnGlnThrValGlyGluProGlyValProGlyProArgGlyLeuPro 370
Db 663 TTCCAGGAGACTTCAGGGACCCCTGGGAGCCTGGGGTGCCTGGACCTCGGGAGACTGCCA 722
QY 371 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProProGlyPro 390
Db 723 GGCTTGCCTGGGCTACAGGATGCCAGGCCCCCAAGGGCCCCCGGGCCCTCTGGCCCA 782
QY 391 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 410
Db 783 TCAGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCCCAACCCCGCCACCGGAGGACAA 842
QY 411 GlyCysProProHisTriPlysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 430
Db 843 GGCTGCCCGCTCACTGGNAGAACTTCACAGACAATGCTACTATTATTTCAGTTGAGAAA 902
QY 431 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 450
Db 903 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGCTTTCACATCTCTGTTTCATA 962
QY 451 AsnThrArgGluGluGlnGlnTriPlysLysGlnMetValGlyArgGluSerHisTriP 470
Db 963 AACACTAGAGAGAAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG 1022
QY 471 IleGlyLeuThrAspSerGluArgGluAsnGluTriPlysTriPlysLeuAspGlyThrSerPro 490
Db 1023 ATCGGCTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCA 1082
QY 491 AspTyrLysAsnTriPlysAlaGlyGlnProAspAsnTriPlysHisGlyHisGlyProGly 510
Db 1083 GACTACAAAAATTTGAAAAGCTGGACAGCCGGATAAATGGGGTCACTGGCCATGGCCAGA 1142
QY 511 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTriPAsnAspPheGluCysGluAspVal 530
Db 1143 GAAGACTGTGCTGGGTGATTATCTGGGCAGTGAAGATTTTCCAATGTGAAGACGTC 1202
QY 531 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 1203 AATACTTCATTTGGCAAAAACAGAGGAGAGACAGTACTGTCTATCTGCATTA 1253
RESULT 10
AC016128/c 169088 bp DNA linear HTG 09-SEP-2000
LOCUS Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 8 unrounded pieces.
AC016128 AC016128.4 GI:10046526
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169088)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-324G2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169088)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Collings,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karacas,A., Klein,J.,
Lehocky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
```

TITLE
JOURNAL
COMMENT
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L999
Center clone name: 324_G-2

Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16522 bases at least Q40
Consensus quality: 167248 bases at least Q30
Consensus quality: 167996 bases at least Q20
Insert size: 177000; agarose-fp
Quality coverage: 168388; sum-of-contigs
Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 17943: contig of 17943 bp in length
* 17944 18043: gap of 100 bp
* 18044 22424: contig of 4381 bp in length
* 22425 22524: gap of 100 bp
* 22525 38094: contig of 15570 bp in length
* 38095 38194: gap of 100 bp
* 38195 49220: contig of 11026 bp in length
* 49221 49320: gap of 100 bp
* 49321 66394: contig of 17074 bp in length
* 66395 66494: gap of 100 bp
* 66495 91692: contig of 25198 bp in length
* 91693 91792: gap of 100 bp
* 91793 120869: contig of 29077 bp in length
* 120870 120969: gap of 100 bp
* 120970 169088: contig of 48119 bp in length.

FEATURES

source

1. 169088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-324G2"
/clone.lib="RPC1-11 Human Male BAC"

misc_feature

1. 17943
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

misc_feature

18044..22424

misc_feature

/note="assembly_fragment"

misc_feature

22525..38094

misc_feature

/note="assembly_fragment"

misc_feature

38195..49220

misc_feature

/note="assembly_fragment"

misc_feature

49321..66394

misc_feature

/note="assembly_fragment"

misc_feature 91793..120869
/note="assembly_fragment"
misc_feature 120970..169088
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN

Alignment Scores:
Pred. No.: 5.17e-34 Length: 169088
Score: 1239.00 Matches: 248
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.73% Indels: 0
DB: 2 Gaps: 0

US-09-763-712A-2 (1-547) x AC016128 (1-169088)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnLeuThrGlnValGln 20
|||||
Db 165612 ATGTATTCTCATTAATGTGTCATCATCAACCTCAACAACCTGACCGAGGTGCAG 165553
Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
|||||
Db 165552 CAGAGGAACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCCAG 165493
Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysAspThr 60
|||||
Db 165492 CGAATCAAGACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGGACACG 165433
Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
|||||
Db 165432 GATTGGCTGAAGAGAAAGTGAGAGCTGCAGACGCTGGCTGCCAACAACCTCTCGGTG 165373
Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
|||||
Db 165372 GCCAAAGCCCAACAGCACACCTGGAGGATATGAACAGCAGCTCAACTCATTCACAGT 165313
Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
|||||
Db 165312 CAGATGGAGAACATCACCACCTATCTCTCAAGCCACGACAGCAACTGAAAGACCTGCAG 165253
Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
|||||
Db 165252 GACTTACACAAAGATGCAGAGAATAGACAGCCATCAAGTTCAACCACTGGAGGAACG 165193
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
|||||
Db 165192 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACAGCCAC 165133
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
|||||
Db 165132 CACCTGGGACCTGACCAGCAATCTAAATGAAGTCAGGACCACTTGCACATACCTT 165073
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
|||||
Db 165072 ACCAAACACACAGATGATCTGACCTCTCTTGAATAATACCTGGCCCAACATCCGTTTGGAT 165013
Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
|||||
Db 165012 TCTGTTTCTCAGGATGCAACAAGATTGTGAGGTCCAGGTTAGACACTGAAGTAGCC 164953
Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
|||||
Db 164952 AACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAGCATGCTCAGCTCATC 164893
Qy 241 LysAsnPheThrIleLeuGlnGly 248
|||||
Db 164892 AAGAATTTTACAATACTACAAGST 164869
RESULT 11
AP000900


```

LOCUS       AP000900                178022 bp    DNA    linear    HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKING
DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION   AP000900
VERSION     AP000900.3  GI:8119043
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-683J11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 178022)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens 178,022 genomic DNA of 18p11.3
JOURNAL     Published Only in Database (1999)
REFERENCE   2 (bases 1 to 178022)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp)
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997403.
            ----- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            ----- Project Information
            Center project name: HumDraft18
            Center clone name: RP11-683J11
            ----- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 161944 bases at least Q40
            Consensus quality: 170357 bases at least Q30
            Consensus quality: 174322 bases at least Q20
            Insert size: 176122; sum-of-contigs
            Quality coverage: 4.56x in Q20 bases; sum-of-contigs
            -----
NOTE: This is a 'working draft' sequence. It currently consists of
20 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
            1 24283: contig of 24283 bp in length
            24384 46875: contig of 22492 bp in length
            46976 65058: contig of 18083 bp in length
            81423 98823: contig of 17401 bp in length
            98924 110068: contig of 11145 bp in length
            110169 122716: contig of 12548 bp in length
            122817 133089: contig of 10273 bp in length
            133190 141206: contig of 8017 bp in length
            141307 149051: contig of 7744 bp in length
            149051 156105: contig of 6955 bp in length
            156105 162048: contig of 5843 bp in length
            162048 162148: contig of 100 bp
            162148 165118: contig of 2970 bp in length
            165118 165218: contig of 100 bp
            165218 167480: contig of 2262 bp in length
            167480 167580: contig of 100 bp
            167580 169986: contig of 2406 bp in length
            169986 170087: contig of 100 bp
            170087 171909: contig of 1823 bp in length
            171909 172009: contig of 100 bp
            172009 172317: contig of 308 bp in length
            172317 172417: contig of 100 bp
            172417 174663: contig of 2246 bp in length
            174663 174764: contig of 100 bp
            174764 176271: contig of 1508 bp in length
            176271 176371: contig of 100 bp
            176371 178022: contig of 1651 bp in length.
FEATURES             Location/Qualifiers
     source            1..178022
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="18"
                        /map="18p11.3"
                        /clone="RP11-683J11"
     misc_feature       1..24283
                        /note="assembly_fragment"
     misc_feature       24384..46875
                        /note="assembly_fragment"
     misc_feature       46976..65058
                        /note="assembly_fragment"
     misc_feature       65159..81322
                        /note="assembly_fragment"
     misc_feature       81423..98823
                        /note="assembly_fragment"
     misc_feature       98924..110068
                        /note="assembly_fragment"
     misc_feature       110169..122716
                        /note="assembly_fragment"
     misc_feature       122817..133089
                        /note="assembly_fragment"
     misc_feature       133190..141206
                        /note="assembly_fragment"
     misc_feature       141307..149050
                        /note="assembly_fragment"

```

* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 24283: contig of 24283 bp in length
 * 24284 24383: gap of 100 bp
 * 24384 46875: contig of 22492 bp in length
 * 46876 46975: gap of 100 bp
 * 46976 65058: contig of 18083 bp in length
 * 65059 65158: gap of 100 bp
 * 65159 81322: contig of 16164 bp in length
 * 81323 81422: gap of 100 bp
 * 81423 98823: contig of 17401 bp in length
 * 98824 98923: gap of 100 bp
 * 98924 110068: contig of 11145 bp in length
 * 110069 110168: gap of 100 bp
 * 110169 122716: contig of 12548 bp in length
 * 122717 122816: gap of 100 bp
 * 122817 133089: contig of 10273 bp in length
 * 133090 133189: gap of 100 bp
 * 133190 141206: contig of 8017 bp in length
 * 141207 141306: gap of 100 bp
 * 141307 149050: contig of 7744 bp in length
 * 149051 149150: gap of 100 bp
 * 149151 156105: contig of 6955 bp in length
 * 156106 156205: gap of 100 bp
 * 156206 162048: contig of 5843 bp in length
 * 162049 162148: gap of 100 bp
 * 162149 165118: contig of 2970 bp in length
 * 165119 165218: gap of 100 bp
 * 165219 167480: contig of 2262 bp in length
 * 167481 167580: gap of 100 bp
 * 167581 169986: contig of 2406 bp in length
 * 169987 170086: gap of 100 bp
 * 170087 171909: contig of 1823 bp in length
 * 171910 172009: gap of 100 bp
 * 172010 172317: contig of 308 bp in length
 * 172318 172417: gap of 100 bp
 * 172418 174663: contig of 2246 bp in length
 * 174664 174763: gap of 100 bp
 * 174764 176271: contig of 1508 bp in length
 * 176272 176371: gap of 100 bp
 * 176372 178022: contig of 1651 bp in length.

FEATURES
 source
 1..178022
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18p11.3"
 /clone="RP11-683J11"
 misc_feature
 1..24283
 /note="assembly_fragment"
 misc_feature
 24384..46875
 /note="assembly_fragment"
 misc_feature
 46976..65058
 /note="assembly_fragment"
 misc_feature
 65159..81322
 /note="assembly_fragment"
 misc_feature
 81423..98823
 /note="assembly_fragment"
 misc_feature
 98924..110068
 /note="assembly_fragment"
 misc_feature
 110169..122716
 /note="assembly_fragment"
 misc_feature
 122817..133089
 /note="assembly_fragment"
 misc_feature
 133190..141206
 /note="assembly_fragment"
 misc_feature
 141307..149050
 /note="assembly_fragment"

Sequence updated (24-Dec-1999).
 Sequence updated (26-May-2000).


```
/note="assembly_fragment"
149151..156105
/note="assembly_fragment"
156206..162048
/note="assembly_fragment"
162149..165118
/note="assembly_fragment clone_end:SP6 vector_side:left"
165219..167480
/note="assembly_fragment"
167581..169986
/note="assembly_fragment"
170087..171909
/note="assembly_fragment"
172010..172317
/note="assembly_fragment clone_end:T7 vector_side:right"
172418..174663
/note="assembly_fragment"
174764..176271
/note="assembly_fragment"
176372..178022
/note="assembly_fragment"
BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others
ORIGIN
Alignment Scores:
Pred. No.: 5.38e-34 Length: 178022
Score: 1239.00 Matches: 248
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.73% Indels: 0
Dbs: 2 Gaps: 0
US-09-763-712A-2 (1-547) x AP000900 (1-178022)
QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuThrGlnValGln 20
Db 13158 ATGATTCCTCAATAGTGTGTCATCATGAACCTCAACACCTGAACCTGACCCAGGPGCAG 13217
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 13218 CAGAGGAACCTCATCAGCAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 13277
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 13278 CGAATCAAGAACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGCACAG 13337
QY 61 AspTyrLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 13338 GATTGGCTCAAGGAGAAAGTCAGAGCTTCAGACGCTGCGTCCCAACACTCTGGCGTTG 13397
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 13398 GCCAAGGCCAACACACACACCCCTGGAGGATATGACAGCCAGCTCAACTCAATTCACAGGT 13457
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluAlaAsnLeuLysAspLeuGln 120
Db 13458 CAGATGGAGAAATCATCACCCTACTCTCTCAAGCCCAAGCAGACAGAACCTGAAAGCCCTGCAG 13517
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 13518 GACTTACACAAAGATGCAGAGAAATGACAGCCATCAAGTTCAACCACTGGAGGAACCC 13577
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 13578 TTCCAGCTCTTTGAGCGGATATTGGAACATCATTAGCAATATCATGTTACACGCCAC 13637
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 13638 CACCTCGGAGCGCTGACCAGCAATCTAAATGAATGAGTCAGGACCACTTGCACGATACCCCT 13697
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 13698 ACCAAACACACAGATGATGTGACCTCTTGAATAATACCTGCGCAACATCCGTTTGGAT 13757
```

```
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 13758 TCTGTTTCTCTCAGGATGCAACAAGATTTCATGAGGTCGAGGTACACTGAAGTAGCC 13817
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeu 240
Db 13818 AACTTATCATGTTATATGGAAGAAATGAAGCTAGTAGTCAAGCATGCTCAGCTCATC 13877
QY 241 LysAsnPheThrIleLeuGlnGly 248
Db 13878 AAGAATTTTACAATACTACTCAAGGT 13901
RESULT 12
LOCUS AP001022 187635 bp DNA linear HTG 30-MAY-2000
DEFINITION DRAFT SEQUENCE, 33 unordered pieces.
ACCESSION AP001022
VERSION AP001022.2 GI:8117692
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-815L4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 187,635 genomic DNA of RP11.3
Published Only in Database (2000)
2 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997772.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft18
Center clone name: RP11-815L4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167192 bases at least Q40
Consensus quality: 177020 bases at least Q30
Consensus quality: 182105 bases at least Q20
Insert size: 184435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
33 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 21332 contig of 21332 bp in length
21433 42072 contig of 20640 bp in length
42173 54336 contig of 12164 bp in length
54437 66766 contig of 12330 bp in length
66867 77849 contig of 10983 bp in length
77950 87798 contig of 9849 bp in length
87899 99510 contig of 11612 bp in length
```

99611 107273 contig of 7663 bp in length
107374 113189 contig of 5816 bp in length
113290 120559 contig of 7270 bp in length
120660 126468 contig of 5809 bp in length
126569 131619 contig of 5051 bp in length
131720 136810 contig of 5091 bp in length
136911 139925 contig of 3015 bp in length
140026 144232 contig of 4207 bp in length
144333 148902 contig of 4570 bp in length
149003 152918 contig of 3916 bp in length
153019 156399 contig of 3381 bp in length
156500 158933 contig of 2434 bp in length
159034 161884 contig of 2851 bp in length
161985 164301 contig of 2317 bp in length
164402 166300 contig of 1899 bp in length
166401 168580 contig of 2180 bp in length
168681 171270 contig of 2590 bp in length
171371 174092 contig of 2722 bp in length
174193 176095 contig of 1903 bp in length
176196 177890 contig of 1695 bp in length
177991 179786 contig of 1796 bp in length
179887 181478 contig of 1592 bp in length
181579 182895 contig of 1317 bp in length
182996 184841 contig of 1317 bp in length
184942 186345 contig of 1404 bp in length
186446 187635 contig of 1190 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 21332: contig of 21332 bp in length
21333 21432: gap of 100 bp
21433 42072: contig of 20640 bp in length
42073 42172: gap of 100 bp
42173 54336: contig of 12164 bp in length
54337 54436: gap of 100 bp
54437 66766: contig of 12330 bp in length
66767 66867: gap of 100 bp
66867 77849: contig of 10983 bp in length
77850 77949: gap of 100 bp
77950 87798: contig of 9849 bp in length
87799 87898: gap of 100 bp
87899 99510: contig of 11612 bp in length
99511 99610: gap of 100 bp
99611 107273: contig of 7663 bp in length
107274 107373: gap of 100 bp
107374 113189: contig of 5816 bp in length
113190 113289: gap of 100 bp
113290 120559: contig of 7270 bp in length
120560 120659: gap of 100 bp
120660 126468: contig of 5809 bp in length
126469 126568: gap of 100 bp
126569 131619: contig of 5051 bp in length
131620 131719: gap of 100 bp
131720 136810: contig of 5091 bp in length
136811 136910: gap of 100 bp
136911 139925: contig of 3015 bp in length
139926 140025: gap of 100 bp
140026 144232: contig of 4207 bp in length
144233 144332: gap of 100 bp
144333 148902: contig of 4570 bp in length
148903 149002: gap of 100 bp
149003 152918: contig of 3916 bp in length
152919 153018: gap of 100 bp
153019 156399: contig of 3381 bp in length
156400 156499: gap of 100 bp
156500 158933: contig of 2434 bp in length
158934 159033: gap of 100 bp

* 159034 161884: contig of 2851 bp in length
* 161885 161984: gap of 100 bp
* 161985 164301: contig of 2317 bp in length
* 164302 164401: gap of 100 bp
* 164402 166300: contig of 1899 bp in length
* 166301 166400: gap of 100 bp
* 166401 168580: contig of 2180 bp in length
* 168581 168680: gap of 100 bp
* 168681 171270: contig of 2590 bp in length
* 171271 171370: gap of 100 bp
* 171371 174092: contig of 2722 bp in length
* 174093 174192: gap of 100 bp
* 174193 176095: contig of 1903 bp in length
* 176096 176195: gap of 100 bp
* 176196 177890: contig of 1695 bp in length
* 177891 177990: gap of 100 bp
* 177991 179786: contig of 1796 bp in length
* 179787 179886: gap of 100 bp
* 179887 181478: contig of 1592 bp in length
* 181479 181578: gap of 100 bp
* 181579 182895: contig of 1317 bp in length
* 182896 182995: gap of 100 bp
* 182996 184841: contig of 1846 bp in length
* 184842 184941: gap of 100 bp
* 184942 186345: contig of 1404 bp in length
* 186346 186445: gap of 100 bp
* 186446 187635: contig of 1190 bp in length.

FEATURES
source

1..187635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="Rp11-815L4"

misc_feature

1..21332
/note="assembly_fragment"
21433..42072
/note="assembly_fragment"
42173..54336
/note="assembly_fragment"
54437..66766
/note="assembly_fragment"
66867..77849
/note="assembly_fragment"
77950..87798
/note="assembly_fragment clone_end:T7 vector_side:right"
87899..99510
/note="assembly_fragment"
99611..107273
/note="assembly_fragment"
107374..113189
/note="assembly_fragment clone_end:SP6 vector_side:left"
113290..120559
/note="assembly_fragment"
120660..126468
/note="assembly_fragment"
126569..131619
/note="assembly_fragment"
131720..136810
/note="assembly_fragment"
136911..139925
/note="assembly_fragment"
140026..144232
/note="assembly_fragment"
144333..148902
/note="assembly_fragment"
149003..152918
/note="assembly_fragment"
153019..156399

Alignment Scores:
Pred. No.:
Score:
5.61e-34 Length: 187635
1239.00 Matches: 248

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.73% Indels: 0
 DB: 2 Gaps: 0

US-09-763-712A-2 (1-547) x AP001022 (1-187635)

Qy 1 MetTySerHisAsnValIleMetAsnLeuAsnLeuThrGlnValGln 20
 Db 58408 ATGTATTCATATCTGTCATCATGAACCTCAACACCTCAACCTGACCGTGCAG 58467
 Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 Db 58468 CAGAGGAACCTCATCACGAATCTGCAGCGGTCTGGGATGACACAAGCAGGCTATCCAG 58527
 Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 58528 CGAATCAGAGACGACTTTCAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACG 58587
 Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 Db 58588 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACCTCTCGCTTG 58647
 Qy 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 58648 GCCAAGGCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAATTCACAGGT 58707
 Qy 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 58708 CAGATGGAGACATCACCATCTCTCTCAGCCACGAGCAGAACCTGAAAGACCTGCAG 58767
 Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 58768 GACTTACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCACCAACTGGAGAACGC 58827
 Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrrThrAlaHis 160
 Db 58828 TTCCAGCTCTTTGACAGCGATATTGTGAACATCATATAGCAATATATGATACACGCCAC 58887
 Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 58888 CACCTGGCGAGCTGACCAAGATCTTAATGAATGATCAGGACCACTTGACAGATACCTT 58947
 Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 58948 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCGTTTGGAT 59007
 Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 59008 TCTGTTCTCTCAGGATGCAACAAGATTGTGATGAGTGCAGGTTAGACACTGAAGTAGCC 59067
 Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 Db 59068 AACTATCAGTGATTATGAGAGAATAGAGCTAGTAGACTCCACGATGGTCAGCTCATC 59127
 Qy 241 LysAsnPheThrIleLeuGlnGly 248
 Db 59128 AAGATTTTACATACTACAAGGT 59151

RESULT 13
 AP000915/c
 LOCUS 188439 bp DNA linear PRI 27-APR-2002
 DEFINITION Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete sequences.
 ACCESSION AP000915
 VERSION AP000915.5 GI:20334314
 KEYWORDS HTG.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Funjiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens genomic DNA
 Published Only in Database (1999)
 2 (bases 1 to 188439)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Funjiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehico-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 On Apr 26, 2002 this sequence version replaced gi:9188470.
 COMMENT
 FEATURES
 Location/Qualifiers
 1. 188439
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18p"
 /clone="RP11-720L2"
 /map="18p"
 BASE COUNT 54003 a 41134 c 40027 g 53275 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5.63e-34 Length: 188439
 Score: 1239.00 Matches: 248
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.73% Indels: 0
 DB: 9 Gaps: 0
 US-09-763-712A-2 (1-547) x AP000915 (1-188439)
 Qy 1 MetTySerHisAsnValIleMetAsnLeuAsnLeuThrGlnValGln 20
 Db 56742 AIGTATTCATATCTGTCATCATGAACCTCAACACCTGACCGTGCAG 56683
 Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 Db 56682 CAGAGGAACCTCATCACGAATCTGCAGCGGTCTGGGATGACACAAGCAGCTATCCAG 56623
 Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 56622 CGAATCAGAGACGACTTTCAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACG 56563
 Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 Db 56562 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACCTCTGCGTTG 56503
 Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 56502 GCCAAGGCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAATTCACAGGT 56443
 Qy 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 56442 CAGATGGAGACATCACCATCTCTCAAGCCCAACGAGCAGAACCTGAAAGACCTGCAG 56383
 Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 56382 GACTTACAAAGATGCAGAGAATAGACAGCCATCAAGTTCACCAACTGGAGGAACGC 56323
 Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrrThrAlaHis 160
 Db 56322 TTCCAGCTCTTTGACAGCGATATTGTGAACATCATAGCAATATCATGTTACAGACCCAC 56263
 Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 56262 CACCTGCGGAGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACAGATACCCCT 56203
 Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 56202 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTGCCAACAATCGCTTTGGAT 56143

```

QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 56142 TCTGTTCTCTCAGGATGCAACAAGATTGATGAGTCTAGACACTGAAGTAGCC 56083
QY 221 AsnLeuSerValIleMetGlnGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 56082 AACTTATCATGATTATGGAAGAATAGAGCTAGTAGACTCCAAAGCATGGTCAGCTCATC 56023
QY 241 LysAsnPheThrIleLeuGlnGly 248
Db 56022 AAGAATTTTACAATACTACAAGST 55999

RESULT 14
AP000939
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION
AP000939
VERSION
AP000939.3 GI:8119080
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens DNA, clone:RP11-839023.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188255)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 188,255 genomic DNA of 18p11.3
Published Only in Database (1999)
2 (bases 1 to 188255)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@isc.riken.go.jp, Tel:81-42-778-9923,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6937751.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@isc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-839023
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171068 bases at least Q40
Consensus quality: 178865 bases at least Q30
Consensus quality: 182445 bases at least Q20
Insert size: 185155; sum-of-contigs
Quality coverage: 4.41x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 16096 contig of 16096 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 51010 contig of 12448 bp in length
75056 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 5817 bp in length
90910 91010 contig of 100 bp
91010 98615 contig of 7606 bp in length
98616 98715: gap of 100 bp
98716 105978: contig of 7263 bp in length
105979 106078: gap of 100 bp
106079 113049: contig of 6971 bp in length
113050 113149: gap of 100 bp
113150 119201: contig of 6052 bp in length
119202 119301: gap of 100 bp
119302 126582: contig of 7281 bp in length
126583 132070: contig of 5388 bp in length
132071 132170: gap of 100 bp
132171 137138: contig of 4968 bp in length
137139 137238: gap of 100 bp
137239 143055: contig of 5817 bp in length
143056 143155: gap of 100 bp
143156 148639: contig of 5484 bp in length
148640 148739: gap of 100 bp
148740 153086: contig of 4347 bp in length
153087 153186: gap of 100 bp
153187 158921: contig of 5735 bp in length
158922 159021: gap of 100 bp
159022 163174: contig of 4153 bp in length
163175 163274: gap of 100 bp
163275 167066: contig of 3792 bp in length

```

98716 105978 contig of 7263 bp in length
 106079 113049 contig of 6971 bp in length
 113150 119201 contig of 6052 bp in length
 119302 126582 contig of 5388 bp in length
 126683 132070 contig of 4968 bp in length
 132171 137138 contig of 5817 bp in length
 137239 143055 contig of 5484 bp in length
 143156 148639 contig of 4347 bp in length
 148740 153086 contig of 5735 bp in length
 153187 158921 contig of 4153 bp in length
 159022 163174 contig of 3792 bp in length
 163275 167066 contig of 2954 bp in length
 167167 170120 contig of 2210 bp in length
 170221 172430 contig of 3068 bp in length
 172531 175598 contig of 1420 bp in length
 175699 177118 contig of 1313 bp in length
 177219 178531 contig of 1226 bp in length
 178632 179857 contig of 1107 bp in length
 179958 181064 contig of 1571 bp in length
 181165 182735 contig of 1088 bp in length
 182836 183923 contig of 1787 bp in length
 184024 185810 contig of 1235 bp in length
 185911 187145 contig of 1010 bp in length
 187246 188255 contig of 1010 bp in length

Sequence updated (06-Jan-2000)
 Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 16096: contig of 16096 bp in length
 16097 16196: gap of 100 bp
 16197 33789: contig of 17593 bp in length
 33790 33889: gap of 100 bp
 33890 51010: contig of 17121 bp in length
 51011 51110: gap of 100 bp
 51111 63558: contig of 12448 bp in length
 63559 63658: gap of 100 bp
 63659 75056: contig of 11398 bp in length
 75057 75156: gap of 100 bp
 75157 84737: contig of 9581 bp in length
 84738 84837: gap of 100 bp
 84838 90909: contig of 6072 bp in length
 90910 91009: gap of 100 bp
 91010 98615: contig of 7606 bp in length
 98616 98715: gap of 100 bp
 98716 105978: contig of 7263 bp in length
 105979 106078: gap of 100 bp
 106079 113049: contig of 6971 bp in length
 113050 113149: gap of 100 bp
 113150 119201: contig of 6052 bp in length
 119202 119301: gap of 100 bp
 119302 126582: contig of 7281 bp in length
 126583 132070: contig of 5388 bp in length
 132071 132170: gap of 100 bp
 132171 137138: contig of 4968 bp in length
 137139 137238: gap of 100 bp
 137239 143055: contig of 5817 bp in length
 143056 143155: gap of 100 bp
 143156 148639: contig of 5484 bp in length
 148640 148739: gap of 100 bp
 148740 153086: contig of 4347 bp in length
 153087 153186: gap of 100 bp
 153187 158921: contig of 5735 bp in length
 158922 159021: gap of 100 bp
 159022 163174: contig of 4153 bp in length
 163175 163274: gap of 100 bp
 163275 167066: contig of 3792 bp in length

* 167067 167166: gap of 100 bp in length
* 167167 170120: contig of 2954 bp in length
* 170121 170220: gap of 100 bp in length
* 170221 172430: contig of 2210 bp in length
* 172431 172530: gap of 100 bp in length
* 172531 175598: contig of 3068 bp in length
* 175599 175698: gap of 100 bp in length
* 175699 177118: contig of 1420 bp in length
* 177119 177218: gap of 100 bp in length
* 177219 178531: contig of 1313 bp in length
* 178532 178631: gap of 100 bp in length
* 178632 179857: contig of 1226 bp in length
* 179858 179957: gap of 100 bp in length
* 179958 181064: contig of 1107 bp in length
* 181065 181164: gap of 100 bp in length
* 181165 182735: contig of 1571 bp in length
* 182736 182835: gap of 100 bp in length
* 182836 183923: contig of 1088 bp in length
* 183924 184023: gap of 100 bp in length
* 184024 185810: contig of 1787 bp in length
* 185811 185910: gap of 100 bp in length
* 185911 187145: contig of 1235 bp in length
* 187146 187245: gap of 100 bp in length
* 187246 188255: contig of 1010 bp in length.

FEATURES

source

1. 188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1. 16096
/note="assembly_fragment"
16197. .33789
/note="assembly_fragment"
33890. .51010
/note="assembly_fragment"
51111. .63558
/note="assembly_fragment"
63659. .75056
/note="assembly_fragment"
75157. .84737
/note="assembly_fragment"
84838. .90909
/note="assembly_fragment"
91010. .98615
/note="assembly_fragment"
98716. .105978
/note="assembly_fragment"
106079. .113049
/note="assembly_fragment"
113150. .119201
/note="assembly_fragment"
119302. .126582
/note="assembly_fragment"
126683. .132070
/note="assembly_fragment"
132171. .137138
/note="assembly_fragment"
137239. .143055
/note="assembly_fragment"
143156. .148639
/note="assembly_fragment"
148740. .153086
/note="assembly_fragment clone_end:SP6 vector_side:right"
153187. .158921
/note="assembly_fragment"
159022. .163174

Alignment Scores:

Pred. No.: 4.87e-33 Length: 188255
Score: 1212.00 Matches: 247
Percent Similarity: 99.60% Conservative: 0

Best Local Similarity: 99.60% Mismatches: 1
Query Match: 40.82% Indels: 1
DB: 2 Gaps: 0

US-09-763-712a-2 (1-547) x AP000939 (1-188255)

QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
|||||
Db 102661 ATGTATTCTCATAATGTGTTCATCATGAACTCAACACCTGAACCTGACCCAGGTGCAG 102720
|||||
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
|||||
Db 102721 CAGAGAACCTCATCAGGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAG 102780
|||||
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
|||||
Db 102781 CGAATCAAGACGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGACACG 102840
|||||
QY 61 AspTyrLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
|||||
Db 102841 GATTGGCTGAAGGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTGCGTTG 102900
|||||
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
|||||
Db 102901 GCCAAAGCCCAACAGCACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 102960
|||||
QY 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
|||||
Db 102961 CAGATGGAGACATCACCACATATCTCAAGCCACAGGACAGACACTGAAAGACCTGCAG 103020
|||||
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
|||||
Db 103021 GACTTACAAAGATGCAGAGAATAGAACAGCCATCAAGTTCAACCACTGGAGGAACGC 103080
|||||
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
|||||
Db 103081 TTCCAGCTCTTTGAGACGGATATTCTGAACATCATAGCAATATCAGTTACACAGCCAC 103140
|||||
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
|||||
Db 103141 CACCTGGGACGCTCACCAGCAATCTAATGAAGTCAGGACCCTTGCACATACCCCTT 103200
|||||
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
|||||
Db 103201 ACCAAACACACAGATGATCTGACCTCCTTGAATAATACCTGGCCAAACATCCGTTT-GAT 103259
|||||
QY 201 SerValSerLeuArgMetClnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
|||||
Db 103260 TCTGTTTCTCTCAGGATGCACAAAGATTGTGAGGTGCGAGGTAGACACTGAAGTAGCC 103319
|||||
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
|||||
Db 103320 RACTTATCAGTGATATGGAGAAATGAAGCTAGTAGACTCCCAAGCATGTCAGCTCATC 103379
|||||

RESULT 15

AC114677/c

LOCUS

DEFINITION

Mus musculus clone RP24-213K19, WORKING DRAFT SEQUENCE, 10 ordered

pieces.

AC114677

VERSION

AC114677.3 GI:21362159

KEYWORDS

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 182029)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Mus musculus, clone RP24-213K19

Db 1877 ATGTATTCTCAGAGCGTGGTTATCATGAAACCTCAACAACCTGAACCTAACCAGGTTCCAG 1818
Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 1817 CAGAGGAACCTTATCTCAAACTCGCAGCAGTCTGTGGATGACACAGGCTGGCCATCCAG 1758
Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
Db 1757 CCAATTAAAGATGATTCCAAAATCTGCAGCAGGTTTCTTCAAGCCAAGAAGCACACC 1698
Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 1697 GATTGGCTAAAGGAAAAAGTACAGACTTGCAGACATTGGCTGCCAACCAACTCTGCCCTG 1638
Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 1637 GCCAAGCCCAACATGACACCTAGAGGATATGATAGCCAGCTCAGCTCATTCACAGGT 1578
Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 1577 CAGATGGACAACATTACCACATCTCACAGGCCAACGAGCAGAGGCTGAAAGACCTTCAG 1518
Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 1517 GACTTACACAAGGATACAGAAAATAAGACAGCTGTCAAGTTTCAGCCAACTTGAGGAAGGC 1458
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 1457 TTCCAGGCTTTGAGACAGATATTGTGAACATCATTAGCAACATCAGCTACACGCCCAT 1398
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1397 CACCTGAGGACACTGACCAGCAATCTGAATGATGTAGGACCACATGCACACACCTTG 1338
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1337 ACCAGACACCGGATGACCTGACCTCCTGTAATAACACACTAGTCAACATCCGCTTGGAT 1278
Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1277 TCTATTCTCTCAGGATGCAGACACATGATGAGTCAAGTTAGACACTGAAGTGCC 1218
Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1217 AACTTATCAGTGGTTATGGAAGAGATGAAACTGGTTGACTCCAGCAGCGTCAGCTCATC 1158
Qy 241 LysAsnPheThrIleLeuGlnGly 248
Db 1157 AGAACTTTACCATTTCTACAAGGT 1134

Search completed: March 21, 2003, 11:12:28
Job time : 4023.46 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:02:25 ; Search time 4292.42 Seconds
(without alignments)
11492.169 Million cell updates/sec

Title: US-09-763-712A-l_COPY_l_1695

Perfect score: 1695

Sequence: 1 gtcacgaatctgcagcaaga.....cagtactgtcatctgcatta 1695

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.Other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.Other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1695	100.0	2983	9	AB005145	AB005145 Homo sapi
2	1693.4	99.9	2641	6	AX047353	AX047353 Sequence
3	1691.8	99.8	3058	9	AB038518	AB038518 Homo sapi
4	1477.6	87.2	2005	6	AX454442	AX454442 Sequence
5	1477.6	87.2	2005	6	AX490920	AX490920 Sequence
6	1359	80.2	2637	10	AB078434	AB078434 Mus muscu
7	1354.2	79.9	3291	10	AB038519	AB038519 Mus muscu
8	1281.2	75.6	4330	9	AB052103	AB052103 Homo sapi
9	1251.4	73.8	1886	9	HS0802985	AL713657 Homo sapi
c 10	798	47.1	169088	2	AC016128	AL713657 Homo sapi
c 11	798	47.1	178022	2	AF000900	AC016128 Homo sapi
c 12	798	47.1	187635	2	AP001022	AP000900 Homo sapi
c 13	798	47.1	188439	9	AP000915	AP001022 Homo sapi
c 14	786	46.4	188255	2	AP000939	AP000915 Homo sapi
c 15	662	39.1	182029	2	AC114677	AP000939 Homo sapi
c 16	662	39.1	193208	2	AC102618	AC114677 Mus muscu
c 17	629.8	37.2	71044	2	AC024368	AC102618 Mus muscu
c 18	437.4	25.8	188255	2	AC102618	AC024368 Homo sapi
c 19	388	22.9	182029	2	AC114677	AP000939 Homo sapi
c 20	388	22.9	193208	2	AC102618	AC114677 Mus muscu
c 21	378.4	22.3	130763	2	AC112416	AC102618 Mus muscu
c 22	248.2	14.6	3636	9	AB007829	AC112416 Rattus no
c 23	248.2	14.6	3685	6	E32511	AB007829 Homo sapi
c 24	248.2	14.6	3810	6	E32509	AB007829 Homo sapi
c 25	222.8	13.1	2215	10	BC026446	E32511 Scavenger r
c 26	146.8	8.7	178022	2	AP000900	E32509 Scavenger r
c 27	146.8	8.7	187635	2	AP001022	BC026446 Mus muscu
c 28	135.6	8.0	1711	9	AB007830	AP000900 Homo sapi
c 29	134.6	7.9	95002	2	AF233390	AP001022 Homo sapi
c 30	134.6	7.9	134514	2	AC025192	AB007830 Homo sapi
c 31	134.6	7.9	157324	9	AC013643	E32510 Scavenger r
c 32	134.6	7.9	175382	9	AF235104	AF233390 Homo sapi
c 33	132.8	7.8	152181	2	AC126444	AC025192 Homo sapi
c 34	132.8	7.8	219809	2	AC125058	AC013643 Homo sapi
c 35	129	7.6	6156	10	AB009993	AF235104 Homo sapi
c 36	129	7.6	2193	3	PALCOL1A	AC126444 Mus muscu
c 37	128.2	7.4	6076	10	AF272661	AB009993 Mus muscu
c 38	125.8	7.3	1041	4	AF138883	M25282 P.lividus a
c 39	124.2	7.3	2053	5	AB008374	AF272661 Rattus no
c 40	124.2	7.3	4502	5	AB052836	AF138883 Bos tauru
c 41	122.6	7.2	4995	10	MMU16789	AB008374 Oncorhync
c 42	122.6	7.2	810	6	AR014116	AB052836 Oncorhync
c 43	121.6	7.2	4628	4	AB008683	U16789 Mus musculu
c 44	121.4	7.2	5676	6	E07265	AR014116 Sequence
c 45	121.4	7.2	5676	6	E07265	AB008683 Bos tauru
						E07265 cDNA encodi

ALIGNMENTS

RESULT 1
AB005145
LOCUS
DEFINITION Homo sapiens-CL-p1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1
KEYWORDS Homo sapiens female tissue; placenta cDNA to mRNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Keshi, H.,
Sakai, Y., Fukuchi, A., Sakamoto, T., Itabe, H., Suzutani, T.,
Ogasawara, M., Yoshida, I. and Wakamiya, N.
PRI 21-NOV-2001

TITLE	The membrane-type collectin CL-P1 is a scavenger receptor on
JOURNAL	J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE	21570232
REFERENCE	2 (bases 1 to 2983)
AUTHORS	Ohtani, K.
TITLE	Direct Submission
JOURNAL	Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College, Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp, Tel: +81-166-68-2393. Fax: +81-166-68-2399)

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

gene

CDS

```

/gene=CD-F1
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"

```

/translation="MKDPAEEVEQSGYRFRQIGTQCTCKNNWALKPSILILY
ILCALUTITVAIGUVKQMDNVTGSKRTQYDKLTAIVEDSKQDQGTGKAI
STNSLSRSLDILROQRELTETSKNKDTLQKSGADSLVQSGQKRETLN
SLITVNTKLVADYITNLOQDTSVLQGNLOQMYSHNVYVMSLNLTOQVRN
LFTNLORSQDTSQAORTKYNQLOQVQKQKTDWIKESVQSLTOAANNSALA
KANNDTLDNLSQNSLTGOMENITLTISQANEQKLDQLDKDAENRTAKENOLE
RFOLFETDILNIISLTSYSLNEVYCTDRLTKHTDDTLNLNNTLANI
RLDSVLSRQQDLMSRLDTEVANUSVLSWEEMKLVCTSRHGLKINLTLOGPPEKGP
GRSGSGPPPTGNGKQGEKGEPPGPAERGEIPGAPGPEERGKSGKSGKPG
GSRGSGKPPQSGKQGEKGPQKQKGPQSGQKQGLQGTQVETGPGYVGRGRLP
LPGVNGKQGPQKGPQSGPAAVLQPNQETPAPDNQSGKPHKWTDKCYFSVE
KEFTEDAKLFCSEDKSHSLYVINTREEQOIKKKQWGRSHIWIQDTERENENKWDJ
TSPDYKNWKAQPDQNGWGHGGPGEDCAGLIYAGQWNPQCEDNFKETCKDRETVLSS
AL"

	AL		
BASE COUNT	914 a	707 c	703 g
ORIGIN			
			659 t

Query Match	100.0%;	Score 1695;	DB 9;	Length 2983;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1695;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	GTCAAGAACTCGACGAAGATACCAAGCGTCTCCAGGCGAATCTGCAGAACCAATGTAT	60		
DB				
602	GTCAAGAACTCGACGAAGATACCAAGCGTCTCCAGGCGAATCTGCAGAACCAATGTAT	661		
QY 61	TCTCATATGTGTGCATCATCAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG	120		
DB				
662	TCTCATATGTGTGCATCATCAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG	721		
QY 121	AACCTCATCAGGATCTGCAGCGGTCTGTGGATGACACACAGCCAGGCTATCCAGCGAATC	180		
DB				
722	AACCTCATCAGGATCTGCAGCGGTCTGTGGATGACACACAGCCAGGCTATCCAGCGAATC	781		
QY 181	AAGAACGACTTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGCACACGATTTGG	240		
DB				
782	AAGAACGACTTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGCACACGATTTGG	841		
QY 241	CTGAAGAGAAAGTGCAGAGCTTGCAGACGGTGGCTGCCAACAACTCTGCGTTTGGCCAAA	300		
DB				
842	CTGAAGAGAAAGTGCAGAGCTTGCAGACGGTGGCTGCCAACAACTCTGCGTTTGGCCAAA	901		
QY 301	GCCAACAACGACACCCCTGGAGGATATCAACAGCCAGCTCAACTCATTCACAGGTCAGATG	360		
DB				
902	GCCAACAACGACACCCCTGGAGGATATGACACGCCAGCTCAACTCATTCACAGGTCAGATG	961		
QY 361	GAGAACAATCACCACTATCTCTCAAGCCAAACGAGCAGAACCTGAAAGACCTTCGACGACTTA	420		
DB				
962	GAGAACAATCACCACTATCTCTCAAGCCAAACGAGCAGAACCTGAAAGACCTTCGACGACTTA	1021		

QY 1501 AGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCGCGGATAAC 1560
 Db 2102 AGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCGCGGATAAC 2161
 QY 1561 TGGGGTCAATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTTATCTGGGCACTGG 1620
 Db 2162 TGGGGTCAATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTTATCTGGGCACTGG 2221
 QY 1621 AACGATTTCCAAATGTGAAGACGTCAATAACTTTCATTTGGCGAAAGACAGGAGACAGTA 1680
 Db 2222 AACGATTTCCAAATGTGAAGACGTCAATAACTTTCATTTGGCGAAAGACAGGAGACAGTA 2281
 QY 1681 CTGTCATCTGCATTA 1695
 Db 2282 CTGTCATCTGCATTA 2296

RESULT 2
 AX047353
 LOCUS AX047353 2641 bp DNA linear PAT 15-DEC-2000
 DEFINITION Sequence 39 from Patent WO0068380.
 ACCESSION AX047353
 VERSION AX047353.1 GI:11876599
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2641)
 AUTHORS Bandman, O., Hillman, J.L., Tang, Y.T., Lal, P., Yue, H., Baughn, M.R.,
 Lu, D.A. and Azimzai, Y.
 TITLE Extracellular matrix and adhesion-associated proteins
 JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
 Incyte Genomics, Inc. (US)

FEATURES
 source Location/Qualifiers
 1..2641
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 4586187CB1"

BASE COUNT 823 a 628 c 608 g 582 t

ORIGIN
 Query Match 99.9%; Score 1693.4; DB 6; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1694; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCACGAATCTGCAGCAAGATACCAGCTGCTCCAGGGCAATCTGCAGAACCAATGTAT 60
 Db 243 GTCACGAATCTGCAGCAAGATACCAGCTGCTCCAGGGCAATCTGCAGAACCAATGTAT 302

QY 61 TCTCATATATGTGTCTATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGACAGAG 120
 Db 303 TCTCATATATGTGTCTATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGACAGAG 362

QY 121 AACCTCATCAGCAATCTGCAGGGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
 Db 363 AACCTCATCAGCAATCTGCAGGGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 422

QY 181 AAGACGACTTTCARAATCTGCAGCAGGTTCCTTCAGGCCAAGAGACACGCGATTGG 240
 Db 423 AAGACGACTTTCARAATCTGCAGCAGGTTCCTTCAGGCCAAGAGACACGCGATTGG 482

QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACAACCTCTGCGTTGGCCAAA 300
 Db 483 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACAACCTCTGCGTTGGCCAAA 542

QY 301 GCCAACACGACACCTCGAGGATATGAACGCCAGCTCACTCACTCATTCACAGGTTCAGATG 360
 Db 543 GCCAACACGACACCTCGAGGATATGAACGCCAGCTCACTCACTCATTCACAGGTTCAGATG 602

QY 361 GAGAACATCACCACATATCTCTCAAGCCACGACGACAGCAACCTGAAGACCTGAGGACTTA 420
 Db 603 GAGAACATCACCACATATCTCTCAAGCCACGACGACAGCAACCTGAAGACCTGAGGACTTA 662

QY 421 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTTCAACCAACTGGAGGAACGCTTCCAG 480
 Db 663 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTTCAACCAACTGGAGGAACGCTTCCAG 722

QY 481 CTCCTTTGACAGCGATATTTGTAACATCATATAGCAATATCAGTTTACACAGCCACGACCTG 540
 Db 723 CTCCTTTGACAGCGATATTTGTAACATCATATAGCAATATCAGTTTACACAGCCACGACCTG 782

QY 541 CGSACCTTGACACAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA 600
 Db 783 CGSACCTTGACACAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA 842

QY 601 CACACAGATGATCTGACCTCTTGAATAATACCTCGCCCAACATCCGTTTGGATTCTGTT 660
 Db 843 CACACAGATGATCTGACCTCTTGAATAATACCTCGCCCAACATCCGTTTGGATTCTGTT 902

QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAAGTAGGCAACTTA 720
 Db 903 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAAGTAGGCAACTTA 962

QY 721 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTTCCAAGCATGCTCAGCTCATCAAGAT 780
 Db 963 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTTCCAAGCATGCTCAGCTCATCAAGAT 1022

QY 781 TTTTACAATACTACAAGGTCCACCGGGCCCCAGGGGTTCCAAGAGGTGACAGAGATCCCAG 840
 Db 1023 TTTTACAATACTACAAGGTCCACCGGGCCCCAGGGGTTCCAAGAGGTGACAGAGATCCCAG 1082

QY 841 GGAACCCCTGGCCCACTGCGCAACAAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCA 900
 Db 1083 GGAACCCCTGGCCCACTGCGCAACAAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCA 1142

QY 901 CTTGGCCCTGCGGGTGCAGAGAGCCCAATTTGGACACAGCTGCTCCCGCCGAGAGCGTGC 960
 Db 1143 CTTGGCCCTGCGGGTGCAGAGAGCCCAATTTGGACACAGCTGCTCCCGCCGAGAGCGTGC 1202

QY 961 GGCACAAAGGATCTAAAGGCTCCACAGGCCCCCAAGAGGTCTCCGTTTCCCTTGGGAAGCCC 1020
 Db 1203 GGCACAAAGGATCTAAAGGCTCCACAGGCCCCCAAGAGGTCTCCGTTTCCCTTGGGAAGCCC 1262

QY 1021 GGCCTCTAGGGCCCCAGTGGGGACCCAGAGCCCCCGGGCCCCCAGGCAAGAGGAGACTC 1080
 Db 1263 GGCCTCTAGGGCCCCAGTGGGGACCCAGAGCCCCCGGGCCCCCAGGCAAGAGGAGACTC 1322

QY 1081 CCGGGCCCTCAGGGCCCTCTGCTGCTTCCAGGGACTTTCAGGGACCTGTTGGGGAGCCCTGG 1140
 Db 1323 CCGGGCCCTCAGGGCCCTCTGCTGCTTCCAGGGACTTTCAGGGACCTGTTGGGGAGCCCTGG 1382

QY 1141 GTGCTCTGGACCTCGGGGACTTGCAGGCTTGCCTTGGGTAGCAGGCTGCGAGGCCCAAG 1200
 Db 1383 GTGCTCTGGACCTCGGGGACTTGCAGGCTTGCCTTGGGTAGCAGGCTGCGAGGCCCAAG 1442

QY 1201 GGGCCCCCGGGCCCTCTGCGCCCATCAGAGCGGTGGTGGCCCTTGGCCCTGCGAGATGAG 1260
 Db 1443 GGGCCCCCGGGCCCTCTGCGCCCATCAGAGCGGTGGTGGCCCTTGGCCCTGCGAGATGAG 1502

QY 1261 CCAACCCCGGACCGGAGGACATGCTGCCCGCCCTCACTTGGAGAACTTTCACAGACAAA 1320
 Db 1503 CCAACCCCGGACCGGAGGACATGCTGCCCGCCCTCACTTGGAGAACTTTCACAGACAAA 1562

QY 1321 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGGCTTTTCTGTGAAGAC 1380
 Db 1563 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGGCTTTTCTGTGAAGAC 1622

QY 1381 AGCTCTTCACATCTTCTTTTCAATAACACTAGAGAGAAACAGCAATGAGATAAACAACAG 1440
 Db 1623 AGCTCTTCACATCTTCTTTTCAATAACACTAGAGAGAAACAGCAATGAGATAAACAACAG 1682

QY 1441 ATGGTAGGAGAGAGAGCCACTTGATCGGCTCACAGACTCAGAGCTGGAATGAATGG 1500
 Db 1683 ATGGTAGGAGAGAGAGAGCCACTTGATCGGCTCACAGACTCAGAGCTGGAATGAATGG 1742

Qy	1501	AAAGTGGCTGGATGGGACATCTCCAGACATACAAAATTTGGAAGCTGGACAGCCCGGATAAC	1560
Db	1743	AAAGTGGCTGGATGGGACATCTCCAGACATACAAAATTTGGAAGCTGGACAGCCCGGATAAC	1802
Qy	1561	TGGGTCATGCGCCATGGCCAGGAGAAAGACTGTCTGGGTTGATTATGTCTGGGCACTGG	1620
Db	1803	TGGGTCATGCGCCATGGCCAGGAGAAAGACTGTCTGGGTTGATTATGTCTGGGCACTGG	1862
Qy	1621	AACGATTTCCTAATGTGAAGAGCTCAATTAACATTCATTATTCGCGAAAAGACAGGAGACAGTA	1680
Db	1863	AACGATTTCCTAATGTGAAGAGCTCAATTAACATTCATTATTCGCGAAAAGACAGGAGACAGTA	1922
Qy	1681	CTGTCACTCGCATTA	1695
Db	1923	CTGTCACTCGCATTA	1937
RESULT 3			
AB038518			
LOCUS			
DEFINITION	AB038518	3058 bp mRNA linear	PRI 08-MAR-2001
ACCESSION	AB038518	type 1, complete cds.	
VERSION	AB038518.1	GI:13365514	
KEYWORDS			
SOURCE			
ORGANISM		Homo sapiens tissue_lib:Placenta cDNA to mRNA.	
REFERENCE		Homo sapiens	
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
JOURNAL		1. (sites)	
MEDLINE		Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F., and Nakamura, T.	
REFERENCE		Molecular cloning and functional characterization of a human	
AUTHORS		scavenger receptor with C-type lectin (SRCL), a novel member of a	
TITLE		scavenger receptor family	
JOURNAL		Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)	
MEDLINE		21092718	
REFERENCE		2 (bases 1 to 3058)	
AUTHORS		Nakamura, K. and Nakamura, T.	
TITLE		Direct Submission	
JOURNAL		Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical	
FEATURES		School, Division of Biochemistry, Biomedical Research Center; 2-2	
source		Yamadaoka, Suita, Osaka 565-0871, Japan	
gene		(E-mail: knakamura@nibch.med.osaka-u.ac.jp)	
CDS		Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)	
		Location/Qualifiers	
		1..3058	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/tissue_lib="Placenta"	
		1..3058	
		/gene="SRCL"	
		209..2437	
		/gene="SRCL"	
		/codon_start=1	
		/product="scavenger receptor with C-type lectin type 1"	
		/protein_id="BAB39147.1"	
		/db_xref="GI:13365515"	
		/translation="MKDDFAEEVEVSFGYKRGIGQGTCTCKNNWALKFSIILY	
		ILCALITVAILGYKVKPMNDVTGSMETSRQYDKLTAVESDLKLDQGTGKAI	
		STNSELSTFRSDLDRLQRLRETEKTSKNKDLLEKIQAGDALVDQROSKLTFENN	
		SFLITVNTKTLQAYNGVTLQDPTVQGNQVSHNVITMNLNLNLTQVQRN	
		LITNLSRVDDTSQALQIKNDFQNLQVFLQAKDLDWLKVKVSLQTLAANNSALA	
		KANNFTLEDNLSQNSFTGQNMETTISQANEQNKLQDLKHDKENRAIKRNFNLEE	
		RFQFLTDIVNIIISNYSYTAHHRLTITSNLNEVRTCTDTLTKHTDLDLSNNLTANI	
		RUDSVLSRHOQDLMSRLDTEAVNLVSIMEEMKLVDKSHGLIKNFITLQGGPPRGP	
		RDRGSGQGGPPTGMRKGEGKEGPPGAPGERGIPGAPGPGRGKSGKSGQPLG	
		LGSVCPMGKPGPGPGSGAVVPLALONEPTAPEDNSCPPEHMKNFDTKCVYFSE	
		KEITFEADKLFCEDKSSHLVFNTRREQQWIKQMGVRESHWIGLTUSERENKWKLDG	
		TSPDYKNRWKAGQPDNNGHGPGCEDCAGLIYAGQWDFQCEDVNNFICKRDRITVLS	
		AL	
		3020..3025	
		/gene="SRCL"	
polyA_signal			

Db 1700 GGCARAAGGATCTAAGGCTCCAGGGCCCCAAAGAGCTCCCTGGTGTCCCTGGGAGGCC 1759
QY 1021 GCCCTCAGGGCCCCAGTGGGACCCAGGCCGCCCGGGCCACCAGGAGGAGGACTC 1080
Db 1760 GGCCCTCAGGGCCCCAGTGGGACCCAGGCCGCCCGGGCCACCAGGAGGAGGACTC 1819
QY 1081 CCGGGCCCTCAGGGCCCCCTCTGGCTTCAGGAGACTTCAGGGACCGTGGGGAGCCTGGG 1140
Db 1820 CCGGGCCCTCAGGGCCCCCTCTGGCTTCAGGAGACTTCAGGGACCGTGGGGAGCCTGGG 1879
QY 1141 GTGCTTGAGACCTCGGGGACTCGAGGCTTGCTGGGTACAGGATCGAGGCCCGCCCAAG 1200
Db 1880 GTGCTTGAGACCTCGGGGACTCGAGGCTTGCTGGGTACAGGATCGAGGCCCGCCCAAG 1939
QY 1201 GGGCCCCCGGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGGCCCTGAGAATGAG 1260
Db 1940 GGGCCCCCGGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGGCCCTGAGAATGAG 1999
QY 1261 CCAACCCCGGACCGGAGGACAATGGTGGCCGCTCTACTGGAAGAACTTCACAGACAAA 1320
Db 2000 CCAACCCCGGACCGGAGGACAATAGCTGCCGCCCTCACTGGAAGAACTTCACAGACAAA 2059
QY 1321 TGTCTACTATTTTCAGTTTCAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGAAGAC 1380
Db 2060 TGTCTACTATTTTCAGTTTCAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGAAGAC 2119
QY 1381 AAGTCTTCACATCTGTGTTTTCATAACACTAGAGAGGAACAGCAATGGATAAAAAACAG 1440
Db 2120 AAGTCTTCACATCTGTGTTTTCATAACACTAGAGAGGAACAGCAATGGATAAAAAACAG 2179
QY 1441 ATGGTAGGAGAGAGGCCACTTGGATCGGCCCTCAGAGCTCAGAGCGTGAATAAATGAATGG 1500
Db 2180 ATGGTAGGAGAGAGGCCACTTGGATCGGCCCTCAGAGCTCAGAGCGTGAATAAATGAATGG 2239
QY 1501 AAGTGGCTGGATGGGACATCTCCAGACTACAAAAATGGAAAGCTGGACGCGGATAC 1560
Db 2240 AAGTGGCTGGATGGGACATCTCCAGACTACAAAAATGGAAAGCTGGACGCGGATAC 2299
QY 1561 TGGGTCTATGGCCATGGGCGAGGAGAACTGTGTGGGTGATTATGCTGGGAGTGG 1620
Db 2300 TGGGTCTATGGCCATGGGCGAGGAGAACTGTGTGGGTGATTATGCTGGGAGTGG 2359
QY 1621 AAGGATTTCCATGTCAGAGACGTCAATTAATCTATTTGGGAAAAACAGAGGAGACAGTA 1680
Db 2360 AAGGATTTCCATGTCAGAGACGTCAATTAATCTATTTGGGAAAAACAGAGGAGACAGTA 2419
QY 1681 CTGTCATCTGCATTA 1695
Db 2420 CTGTCATCTGCATTA 2434

RESULT 4
LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 27 from Patent WO20028284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Baker, K.P., Ferrara, N., Gerber, H., Gerritsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.

TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,
Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;

Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
I. (US)

FEATURES Location/Qualifiers
Source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN

Query Match 87.2%; Score 1477.6; DB 6; Length 2005;
Best local Similarity 99.7%; Pred. No. 0;
Matches 1491; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 200 TGCAGCAGGTTTCTTCAAGCAAGAGGACACGATGCTGCTGAAGGAGAAAGTGCACA 259
Db 9 TCCGGCAGGTTTCTTCAAGCAAGAGGACACGATGCTGCTGAAGGAGAAAGTGCACA 68
QY 260 GCTTCAGACGCTGGCTGCCAAACTCTGCTTGGCCAAAGCCAAACACACCCCTGG 319
Db 69 GCTTCAGACACTGGCTGCCAAACTCTGCTTGGCCAAAGCCAAACACACCCCTGG 128
QY 320 AGGATATGAACAGCAGCTCAACTATTTCACAGGTCAGATGGAGACATCACCCTATCT 379
Db 129 AGGATATGAACAGCAGCTCAACTATTTCACAGGTCAGATGGAGACATCACCCTATCT 188
QY 380 CTCAGGCCAACAGCAGAACCTGAAAGCCTGCAGACTTACACAAAGATGAGAGAATA 439
Db 189 CTCAGGCCAACAGCAGAACCTGAAAGCCTGCAGACTTACACAAAGATGAGAGAATA 248
QY 440 GAACGCCATCAAGTTCAACCAACTGGGAAAGCTTCCAGCTCTTTGAGAGGGATATTG 499
Db 249 GAACGCCATCAAGTTCAACCAACTGGGAAAGCTTCCAGCTCTTTGAGAGGGATATTG 308
QY 500 TGAACATCATTAAGCAATATCAGTTTACAGAGCCACACCTCGGAGCGTGCAGCAATC 559
Db 309 TGAACATCATTAAGCAATATCAGTTTACAGAGCCACACCTCGGAGCGTGCAGCAATC 368
QY 560 TAAATGAAGTCAGGACCACTTCACAGATACCTTACCACAAACACAGATGATCGACCT 619
Db 369 TAAATGAAGTCAGGACCACTTCACAGATACCTTACCACAAACACAGATGATCGACCT 427
QY 620 CCTTGAATAAATACCTGGCCAAACATCCGTTTGGATTCTCTCTCAGGATGCAACAAG 679
Db 428 CCTTGAATAAATACCTGGCCAAACATCCGTTTGGATTCTCTCTCAGGATGCAACAAG 487
QY 680 ATTTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATATGGAAGAA 739
Db 488 ATTTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATATGGAAGAA 547
QY 740 TGAAGCTAGTACCTCCAAAGCATGGTCAGCTCATCAAGAATTTTACAATACTACAAGGTC 799
Db 548 TGAAGCTAGTACCTCCAAAGCATGGTCAGCTCATCAAGAATTTTACAATACTACAAGGTC 607
QY 800 CACCGGGCCCCAGGGTCCAAAGAGTGCAGAGGATCCAGGGAGCCCTGGCCCCCAACTG 859
Db 608 CACCGGGCCCCAGGGTCCAAAGAGTGCAGAGGATCCAGGGAGCCCTGGCCCCCAACTG 667
QY 860 GCACAAAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAGA 919
Db 668 GCACAAAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAGA 727
QY 920 GAGGCCCAATTCAGCAGCTGCTCCCGGGAGAGCGTGGCGGCAAGGATCTAAAGGCT 979
Db 728 GAGGCCCAATTCAGCAGCTGCTCCCGGGAGAGCGTGGCGGCAAGGATCTAAAGGCT 787
QY 980 CCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAGGGCCCCAGTG 1039
Db 788 CCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAGGGCCCCAGTG 847
QY 1040 GGGACCCAGGGCCCCCGGGGCCCAACAGGAGGACTCCCGGCCCTCAGGGGCCCTC 1099
Db 1040 GGGACCCAGGGCCCCCGGGGCCCAACAGGAGGACTCCCGGCCCTCAGGGGCCCTC 1099

Db	848	GGGACCCAGGCCCCCGGGGCCACACAGGACCAAGAGGGAACTCCCCCGGGCCCTCAGGGCCCTC	907
Qy	1100	CTGGCTTCCAGGGACTTCAGGGCCACCGCTTGGGGAGCCTCTGGGCTGCCTGGACCTCGGGAC	1159
Db	908	CTGGCTTCCAGGGACTTCAGGGCCACCGCTTGGGGAGCCTCTGGGCTGCCTGGACCTCGGGAC	967
Qy	1160	TGCCAGGCTTCCCTGGGGTACAGGCATGCGCAGGCCCAAGGGCCCCCGGGCCCTCCTG	1219
Db	968	TGCCAGGCTTCCCTGGGGTACAGGCATGCGCAGGCCCAAGGGCCCCCGGGCCCTCCTG	1027
Qy	1220	GCCCATCAGGAGCGGTGGTGGCCCTGCGCCCTGCAGAAATGAGCCCAACCCCGCACCGGAGG	1279
Db	1028	GCCCATCAGGAGCGGTGGTGGCCCTGCGCCCTGCAGAAATGAGCCCAACCCCGCACCGGAGG	1087
Qy	1280	ACAATGGCTGCCCGCCTCACCTGGAAGAACTTCACAGACAAAATGCTACTATTTTTCAGTTG	1339
Db	1088	ACAATGGCTGCCCGCCTCACCTGGAAGAACTTCACAGACAAAATGCTACTATTTTTCAGTTG	1147
Qy	1340	AGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTT	1399
Db	1148	AGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTT	1207
Qy	1400	TCATAAACACTAGAGAGAACGCAATGGATTAATAAACAACACATGGTAGGGAGAGAGACC	1459
Db	1208	TCATAAACACTAGAGAGAACGCAATGGATTAATAAACAACACATGGTAGGGAGAGAGACC	1267
Qy	1460	ACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGAATGAATGAATGAATGAAT	1519
Db	1268	ACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGAATGAATGAATGAATGAAT	1327
Qy	1520	CTCCAGACTCAAAAATTTGGAAGCTGGACAGCCGGATAACTTGGGGTCAATGGCCATGGGC	1579
Db	1328	CTCCAGACTCAAAAATTTGGAAGCTGGACAGCCGGATAACTTGGGGTCAATGGCCATGGGC	1387
Qy	1580	CAGGAGAGACTGCTGGGTGATTTATGCTGGCAGTGAACGATTTCACATGTGAAG	1639
Db	1388	CAGGAGAGACTGCTGGGTGATTTATGCTGGCAGTGAACGATTTCACATGTGAAG	1447
Qy	1640	ACGTCAATTAATTTTCGAAAAAGACAGGAGACAGTACTGTCACTGCATTA	1695
Db	1448	ACGTCAATTAATTTTCGAAAAAGACAGGAGACAGTACTGTCACTGCATTA	1503
RESULT 5			
LOCUS	AX490920	2005 bp	DNA linear PAT 16-AUG-2002
DEFINITION	Sequence 27 from Patent WO0200690.		
ACCESSION	AX490920		
VERSION	AX490920.1	GI:22323797	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1		
AUTHORS	Baker, K.P., Ferrara, N., Gerber, H., Gerttsen, M.E., Goddard, A., Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J., Paponi, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I. and Ye, W.		
TITLE	Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis		
JOURNAL	Patent: WO 0200690-A 27 03-JAN-2002;		
FEATURES	Genentech, Inc. (US)		
source	Location/Qualifiers		
	1..2005		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	606 a	493 c	415 g
ORIGIN			
Query Match	87.28;	Score 1477.6;	DB 6; Length 2005;
Best Local Similarity	99.7%;	Pred. No. 0;	
Matches 1491; Conservative	0;	Mismatches	4; Indels 1; Gaps 1;


```
QY 1280 ACAATGGCTGCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTG 1339
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1088 ACAATAGCTGCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTG 1147
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1340 AGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTGTTT 1399
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1148 AGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTGTTT 1207
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1400 TCATAAAGCACTAGAGAGAACAGCAATGATATAAAAAACAGATGTTAGGAGAGAGAGCC 1459
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1208 TCATAAAGCACTAGAGAGAACAGCAATGATATAAAAAACAGATGTTAGGAGAGAGAGCC 1267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1460 ACTGATCGGCTCACAGACTCAGAGCTGAAATGAATGGAAGTGGCTGGATGGACAT 1519
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1268 ACTGATCGGCTCACAGACTCAGAGCTGAAATGAATGGAAGTGGCTGGATGGACAT 1327
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1520 CTCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGTATGCCATGGCC 1579
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1328 CTCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGTATGCCATGGCC 1387
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1580 CAGGAGAAAGCTGTCTGGGTGATTATGCTGGGCACTGGAACGATTTCCAAATGTGAAG 1639
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1388 CAGGAGAAAGCTGTCTGGGTGATTATGCTGGGCACTGGAACGATTTCCAAATGTGAAG 1447
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1640 ACGTCAATTAATTTGCGAAAAGACAGGAGACAGTACTGTCAATCTGCATTA 1695
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1448 ACGTCAATTAATTTGCGAAAAGACAGGAGACAGTACTGTCAATCTGCATTA 1503
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 6
AB078434 2637 bp mRNA linear ROD 18-JUL-2002
LOCUS Mus musculus CL-P1 mRNA for collectin placenta 1, complete cds.
DEFINITION AB078434
ACCESSION AB078434
VERSION AB078434.1 GI:21901968
KEYWORDS Mus musculus tissue_lib:Liver cdna to mRNA.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuoh,A., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.
cDNA cloning of mouse CL-P1 gene
Unpublished
2 (bases 1 to 2637)
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuoh,A., Sakamoto,T. and Wakamiya,N.
Direct Submission
TITLE Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College,
JOURNAL Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido
AUTHORS 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp,
Tel:81-166-68-2393, Fax:81-166-68-2399)
FEATURES
source
1..2637
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_lib="Liver"
1..2637
/gene="CL-P1"
92..2320
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/db_xref="GI:21901969"
/translation="MKDDFAEEVEVSQYKRFQIEGCTQCTCKNNWALKFSLVILY
ILCALLITVAILGYKVKVKNMDVNTDGMTSHQYDNKLTAVESDLKLGDOAGKKAL
STNSLSTFRSDILDLQOLQEIETKTSKNKDTLEKLAQNGDSLVDROSQLEKTLQNN
SFLITVKNKTLQANGYVYNLQDTSVLQGNLQSNQYSQSVVIMNINLNTVQQRN
LISNLQSVDDTSIARIKNDNFQNFQIQAQKTDLWLEKRVDSQTLAANNSALA
KANNDLTLEDMSQLSSFTGQMDNITTISQANEQSLKDLQDLHKDLENRTAVKFSQLEE
```

```
RFQVETDIVNIISNYSYTAHLRLTSLNLDVTRTTCTDITL RHTDDLTSLNLTINI
RLDSLSLRMOODMMRSKLDTEVANLSVYMEEMKLVDSKHGOLIKFTLQGPGRGP
KDRGSGPGPTKNGOKGEKPGPGPAGRGCTIGVPGPGRSGKSGKSGQGP
GSRSPKPGPGSGDGPFGPKDGLPGPQGPQGLQCTVGEPCVPGPGLPG
LPVPGMPGPGPPGSGPSPGMEPLALONEPTPASEVNGPPHMKNTDKCYFSL
KEITFDKLFCEKSSHLVFINREEOQWIKHTVGRSHWIGLTDSEKMWLDG
SPVDYKNWKAGQPDNWSGHPGDCAGLIYAGOWNDFQCDIEINNFICEKREAVPSS
IL"
BASE COUNT 829 a 617 c 635 g 556 t
ORIGIN
Query Match 80.2%; Score 1359; DB 10; Length 2637;
Best Local Similarity 87.6%; Pred. No. 9e-310;
Matches 1485; Conservative 0; Mismatches 210; Indels 0; Gaps 0;
QY 1 CTCAGAAATCTGCAGCAAGATACCAGCGTCTCCAGGCGCAATCTCAGAACCAATGTAT 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 623 CTCAGAAATCTGCAGCAAGATACTAGTGTCTCCAGGCGCAATCTCAGAGCCAAATGTAT 682
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 61 TCTCAATATGTGTGCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 683 TCTCAGAGCGTGTATCATGAACCTCAACAACCTGAACCTAACCCAGGTTCAGCAGAGG 742
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 121 AACCTCATCACGAATCTGCAGCGGTCTGTGGATGACACAAGCCAGGCTTCCAGCGAATC 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 743 AACCTTATCTCAAAATCTGCAGCAGTCTGTGGATGACACAAGCTTGCCCATCCAGCGAAT 802
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 181 AGAAGCAGCTTTCAAAATCTGCAGCAGGTTTTTCTCAAGCCAAAGACACAGGATTCG 240
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 803 AAGAATGATTTCCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGACACACCGATTCG 862
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 241 CTGAAGGAGAAAGTCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTCGCTTGGCCAAA 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 863 CTAAAGGAAAAGTACAGAGCTTGCAGACATTTGGCTGCCAACAACTCTCGCTTGGCCAAA 922
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 301 GCCAACACGACACCCCTGGAGGATATGAACGCCAGCTCAACTCATTTACAGGTTCAGATG 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 923 GCCAACAAATGACACCCCTAGAGGATATGAATGCCAGCTCAGCTCATTTACAGGTTCAGATG 982
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 361 GAGAATCATCACCATCTCTCTCAAGCCACGAGCAGAACCTTGAAGACCTGGAGGACTTAA 420
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 983 GACAACATTACCATCTATCTCACAGGCCAACGAGCAGAGCGCTGAAGACCTTCAGGACTTAA 1042
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 421 CACAAAGATGCAGAGAAATAGAACGCCATCAAGTTCAACCAACTGGAGAACGCTTCCAG 480
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1043 CACAAAGATACAGAAAATAGAACGCTGCAAGTTCAAGCCAACTTCAGGAACGCTTCCAG 1102
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 481 CTCCTTTGAGACGGATATTTGTGAACATCATTTAGCAATATCAGTTACACGCCCCACCTCG 540
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1103 GTCTTTGAGACAGATATTGTGAACATCATTTAGCAACATCAGCTACAGCCCCATCAGCTG 1162
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 600
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1163 AGGACACTGACCAAGCAATCTGAATGATGTAGACCACATGACACACACTTGGACCAGA 1222
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 601 CACACAGATGATCTGACCTCTTGAATAATACCCCTGGCCAACTCCGTTTGGATTCCTGT 660
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1223 CACACGGATGACCTGACCTCTTGAATAACACACTAGTCAACATCGCTTGGATTCCTATT 1282
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 661 TCTCTCAGATGCACAGATTTTATGAGTTCGAGTTAGAGCTTACAGCTGAAGTACGCAACTTA 720
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1283 TCTCTCAGATGCACAGACATGATGAGTCAAAAGTTAGACACTGAAGTGGCCAACTTAA 1342
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAT 780
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1343 TCAGTGATTATGGAAGAGATGAAGCTAGTTGACTCCAAAGCGGTGAGCTCATCAAGAAC 1402
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 781 TTTACAATACTACAAGGTCCACCGGCCCGCCAGGGGTCCAAAGAGGTGACAGAGATCCCAAG 840
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1403 TTTACCATTCTACAGGTCTCTCGCCCGCAGAGGTCCAAAGGTGACAGAGAGATCTCAG 1462
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 841 GGACCCCTCGCCCAACTGGCAACAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```


Db	1463	GGACCACCTGGTCTCAACTGGCAACAAGGGACAGAAAAGGAGAGAGAGGAGACGCTGGTGCCA	1522
Qy	901	CCTGGCCCTCGGGGTGAGAGAGAGCCCAATTTGGACACAGCTGTGTCCCCCGGAGAGCGTGGC	960
Db	1523	CCTGGCCCTCGGGGTGAGAGAGGGGCACAATTTGGACACAGCTGGCCCTCTCTGGAGAGCGTGGC	1582
Qy	961	GGCAAGAGATCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAAGCC	1020
Db	1583	AGCAAGAGATCCAAGGCTCACAGGGTCCCAAGAGATCTCGTGGTCCCCCAGGGAAGCCT	1642
Qy	1021	GGCCCTCAGGGCCCCAGTGGGACCCAGGCCCCCGGGGCCACACAGGCAAGAGAGGACTC	1080
Db	1643	GGCCCTCAAGGACCTAGTGGGACCCAGGACCAACAGTCCACAGGCAAGATGGACTC	1702
Qy	1081	CCGGCCCTCAGGGCCCTCTCGGCTTCAGGGACTTCAGGGACCTTCAGGGACCGTGGGGAGCCCTGGG	1140
Db	1703	CCTGGCCCTCAGGGCCCTCTCGGCTTCAGGGACTACAGGGACTGTGGGTGAGCCCTGGGA	1762
Qy	1141	GTGCTTGACCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCCCAAG	1200
Db	1763	GTACCTGACCTCGGGGTGGCCAGGCTTGCCAGGGGTGCCAGGCATGCCCTGGGCCCTAAG	1822
Qy	1201	GGCCCCCGGCCCTCTCTGGCCCATCAGGAGCGGTGTGGCCCTGGGCCCTGCAGAAATGAG	1260
Db	1823	GGACACCTGGCCCTCCAGGCCCCCTCAGGAGCAATGGAGCCATTGGCTCTGCAGAAATGAA	1882
Qy	1261	CCAAACCCGGCAGGAGGACAATGGCTGCCGCCCTCAGTGGAGAGACTTCACAGACAAA	1320
Db	1883	CCAAACCCAGCATCAGAGGTCAACGGATGTCGGCCTCAGTGGAGAGACTTCACAGATAAA	1942
Qy	1321	TGCTACTATTTTTTCAGTTGAGAAAATAATTTTTGAGGATGCAAAAGCTTTTCTGTGAAGAC	1380
Db	1943	TGCTACTATTTTTTCATGGAAAATAATTTTTCAAGATGCTAAGCTTTTCTGTGAAGAC	2002
Qy	1381	AAGTCTTCACATCTGTTTTCATTAAGACATAGAGAGAACAGCAATGGATAAAAAACAG	1440
Db	2003	AAATCTTCCCATCTCGTTTTTCATTAATCAAGAGAAGAACAGCAATGGATAAAAAAGCAT	2062
Qy	1441	ATGTAGGGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAAATGAATGG	1500
Db	2063	ACGTGGGGAGAGAAAGCCATTGGATCGGCTCACAGACTCAGAAACAGAAAGCAATGG	2122
Qy	1501	AAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCCGGATAAC	1560
Db	2123	AAGTGGCTAGACGGGTCACTGTTGATTACAAAATACTGGAAAGCTGGACACACAGATAAC	2182
Qy	1561	TGGGCTCATGCCATGGCCAGAGAGACATGTCTGGGTGATTATATGTGGGCGAGTGG	1620
Db	2183	TGGGCGCATGGCCATGGCCAGGAGAGACATGTCTGGCTTGATTATACGAGGACAGTGG	2242
Qy	1621	AACGATTTCGAATGTGAAGAGTCAATAACTTCATTTCCGAAAAGACAGGGAGACAGTA	1680
Db	2243	ATGACTTCCAGTGTGATGAATCAATAACTTCATTTGTGAGAAGAAAGGGAGGCGAGTA	2302
Qy	1681	CTGTCAATCTGCATTA	1695
Db	2303	CCATCATCCATATTA	2317
RESULT 7			
AB038519			
LOCUS	AB038519	3291 bp mRNA linear	ROD 18-JAN-2002
DEFINITION	Mus musculus srl mRNA for scavenger receptor with C-type lectin, complete cds.		
ACCESSION	AB038519		
VERSION	AB038519.1	GI:18146951	
KEYWORDS			
SOURCE	Mus musculus cDNA to mRNA.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T.		

TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin (SRCR)(1), a novel member of the scavenger receptor family
JOURNAL Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
MEDLINE 21575692
REFERENCE 2 (bases 1 to 3291)
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical School, Division of Biochemistry, Biomedical Research Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
FEATURES (E-mail:knakamura@nibch.med.osaka-u.ac.jp, Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
source Location/Qualifiers
1..3291
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..3291
/gene="srcr"
CDS 77..2305
/gene="srcr"
/codon_start=1
/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDDFAEEVEVSFGYKRFGEHGTQCTKINNWLKFSIVLLYILCATITVAIIIGYKVKBMNDVSDMETSHTQYDNKLTAVESDLKLGDQAGKKALSTNSELSTFRSLLDLROOLETETKTSKNKDTLEKLAQSDLSVDROSOLAKTLQNN SFUITYVNTLQAYNGYYVNLQDTHVNLQGNLQSNYSQSVYIMNLNUNLTVQVQRN LLSNLQSDVDTSLAQRKNQFONQQLQVFLQAKDLDWLKVKVSLQVTLAANNLSALA KANNLTDLMNSLSLSTGQMONTTISQANEQSLQDLQHLKHDENTRTAVKFSOLEE RFQVETITVITNSYTAHLRTLTLNSLNDVMTTCTDLTRHDDTLTSLNNTLVNI RLDISILRMOODMRSKLDTDEVANLSVYMEEMKLVDSKHQLIKNFTILQGPKPRGP KCDRGSQGGPGTNGKQGEKGGPGGPGAGERTIGVPGPGERGSGKSGSGQPG LSGRSGPKGPGQPSDGGPGGPGKDGILPQGGPGFQGLQGTGVEGPGVPPRGLPG LSPVGMPPGPGPGGPGGPPGALQNEPTPASEVNGCPPHKKNFYDKCYFSLFE KETLEDAKFLCEDKSSHLVFINRSREQWIKKHYGRSHWIGLGLTDSQESKMWILDG SPVDYKKNWAGQPDNNGSGHGPGEDCAGLIYAGQWNDFQCDIENNFICSEKREAVPSS IL"
polya_signal 3246..3251
/gene="srcr"
BASE COUNT 1013 a 757 c 759 g 762 t
ORIGIN
Query Match 79.9%; Score 1354.2; DB 10; Length 3291;
Best Local Similarity 87.4%; Pred. No. 1.3e-308;
Matches 1482; Conservative 0; Mismatches 213; Indels 0; Gaps 0;
Qy 1 GTCACGAATCTGCAGCAAGATACCAAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db 608 GTCAAAATCTGCACCAAGATACTAATGTCTCCAGGGCAATCTGCAGAGCCAAATGTAT 667
Qy 61 TCTCATATGTGTCATCATGAACCTCAACAACCTGAACCTGACCAGGTGCAGCAGAGG 120
Db 668 TCTCAGACGCTGGTATCATGAACCTCAACAACCTGAACCTAACCCAGGTTCAGCAGAGG 727
Qy 121 AACCTCATCAGGAATCTGCAGCGGTCTGTGGATGACACAAGCCAGGCTATCCACGGAATC 180
Db 728 AACCTTATCTCAATCTGCAGCAGTCTGTGGATGACACAAGCCTGGCCATCCAGCAATT 787
Qy 181 AAGAACGACTTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGACACGGATTGG 240
Db 788 AAGAATGATTCCAAAATCTGCAGCAGGTTTTCCTTCAAGCCCAAGAGACACCGATTGG 847
Qy 241 CTGAAGCAGAAATGTCAGAGCTTGCACACGCTGGCTGCCAACCAACTCTGCGTTGGCCAAA 300
Db 848 CTAAGAGAAAAGTACAGAGCTTGCAGACATATGGCTGCCAACCAACTCTGCGCTGGCCAAA 907
Qy 301 GCCAACAGCAGACCCCTGGAGGATATCAACAGCCAGCTCAACTCATTCACAGGTCAGATG 360
Db 908 GCCAACATGACCCCTAGAGGATATCAATAGCCAGCTCAGCTCATTCACAGGTCAGATG 967
Qy 361 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGAACCTGAAGACACCTGCAGGACTTA 420
Db 967 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGAACCTGAAGACACCTGCAGGACTTA 420

```
Db 968 GACACATTACACATATCTTCACAGGCCAACAGCAGAGAGCCTGAAAGACCTTCAGGACTTA 1027
Qy 421 CACAAAGATGACGAAGATAGAACAGCCATCAAGTTCAACCACTGGAGAACGCTTCACAG 480
Db 1028 CACAAAGGATACAGAAAATAGAACAGCTGTCAAGTTACGCCAACTTGAAGAACGCTTCACAG 1087
Qy 481 CTCCTTTGAGACGGATATTTGTGAACATCATATTGCAATATCAGTATACAGACCCACACCTG 540
Db 1088 GTCTTTTGAGACAGATATTTGTGAACATCATATTGCAACATCAGCTACACAGCCCATCACCTG 1147
Qy 541 CGGACGCTGACCAGCAATCTTAATGAAGTCAGGACCCACTTGCACAGATACCCCTTACCAAA 600
Db 1148 AGACACTGACACCAACATCTGAATGATGTTGGACCAATGACACAGACACCTTGAACCA 1207
Qy 601 CACACAGATGATCTGACCTCTTGAATATACCTTGGCCAAACATCCGTTTGATCTCTGTT 660
Db 1208 CACACGGATGACCTGACCTCTTGAATTAACACACTAGTCAACATCCGCTTGGATCTATT 1267
Qy 661 TCTCTCAGGATGCAACAGATTTTGATGAGGTCGAGGTTAGACACATGAAGTAGCCAACTTA 720
Db 1268 TCTCTCAGGATGCAACAGATGATGAGGTCGAGGTTAGACACTGAAGTGCCCAACTTA 1327
Qy 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAT 780
Db 1328 TCAGTGATTATGGAAGAGATGAACCTGTTGACTCCAAGCAGGTCAGCTCATCAAGAAC 1387
Qy 781 TTTACAATACTACAAGGTCCACCGGCCCCCAGGGGTCCAAGAGGTGACAGAGATCCCAAG 840
Db 1388 TTTACCAATTTCAAGGTCTCTCTGCCCCCAGAGGTCCAAGAGGTGACAGAGATCTCAG 1447
Qy 841 GGACCCCTTGGCCCAACTGGCAACAGGACAGAAAGAGAGAGAAAGGGGAGCCTGGACCA 900
Db 1448 GGACCACTGTGTCACACTGGCAACAGGACAGAAAGAGAGAGAAAGGGGAGCCTGGTCCA 1507
Qy 901 CCTGCGCTTGGCGGTGAGAGAGGCCAATTGGACACAGCTGGTCCCGCGGAGAGCGTGGC 960
Db 1508 CTTGCGCTTGGCGGTGAGAGGGGCAATTTGACCACTGCGGCCCTTCTGGAGAGCGTGGC 1567
Qy 961 GGCAAGGATCTAAAGGCTCCAGGGCCCCCAAGGCTCCCGTGGTTCCCTTGGGAAGCCC 1020
Db 1568 AGCAAGGATCCAAAGGCTCACAGGTCCTCAAGAGTCTCGTGGTCCCGCAGGAAGCT 1627
Qy 1021 GCGCTCAGGCCCCCAGTGGGACCCAGGCCCGCCCGGCCACACAGGCAAGAGGAGCTC 1080
Db 1628 GCGCTCAGGACCTAGTGGGACCCAGGACCCAGGTCACAGGTCACAGGCAAGGATGACTC 1687
Qy 1081 CCGCGCCCTCAGGGCCCTCTGCTTCCAGGACCTTCAGGGCACCGTTGGGAGCGCTGGG 1140
Db 1688 CTTGCGCTCAGGGCCCTCTGCTTCCAGGACCTTCAGGGCACCTGCTGGTGGAGCTGGA 1747
Qy 1141 GTGCTTGACCTCGGGGACTGCGAGGCTTGGCGGTACAGGCAATCCAGGCCCAAG 1200
Db 1748 GTACCTGGACCTCGGGGGTTGCGAGGCTTGGCAGGGGTGCCAGGCAATGCTGGGCTAAG 1807
Qy 1201 GCGCCCCCGGCTCTGCGCCCATCAGAGCGGTGTCGCCCTGCGCCCTGCGAGATGAG 1260
Db 1808 GGACCACTTGGCCCTTCCAGGCCCCCAGGCAATGGAGCCATTTGGCTGTCAGAAATGAA 1867
Qy 1261 CCAACCCCGGACCGGAGCAATGGCTGCGCCGCTCACTGGAAGAACTTTCACAGACAAA 1320
Db 1868 CCAACCCCGCATCAGAGTCAACGGATGTCGCCCTCACTGGAAGAACTTTCACAGATAA 1927
Qy 1321 TGCTACTATTTTTCAGTTGAGAAAATAATTTTGGAGGATGCAAGCTTTCTGTGGAAGAC 1380
Db 1928 TGCTACTATTTTTCAGTTGAGAAAATAATTTTGGAGATGCTTAAGCTTTCTGTGGAAGAC 1987
Qy 1381 AAGTCTTCACATCTGTTTTCATAAACAACACTAGAGAGAACACAAATGGAATAAAAAACAG 1440
Db 1988 AAATCTTCCACTCTGTTTTCATAAACAACACTAGAGAGAACACAAATGGAATAAAAAACAT 2047
Qy 1441 ATGGTAGGAGAGAGAGCCACTGGATCGGCCTCAGACTCAGAGCGTGAATGAATGG 1500
Db 2048 ACCGTGGGAGAGAAAGCCATTGGATCGGCCTCAGACTCAGAGCGTGAATGAATGG 2107
```

```
Qy 1501 AAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAAGCTGGACAGCCGGATAC 1560.
Db 2108 AAGTGGCTAGACGGCTCACCTGTGATTACAAAAAGCTGGAAAGCTGGACCAACAGATAC 2167
Qy 1561 TGGGTCATGTCAGCCATGGCCAGGAGAACTGTGCTGGTTGATTATCTGCTGGCAGTGG 1620
Db 2168 TGGGTCAGTGGCCATGGCCAGGAGAACTGTGCTGGTTGATTATCTGCTGGCAGTGG 2227
Qy 1621 AACGATTTCCAAATGTGAAGACGTCAATAACTTCAATTTGGGAAAAAGACAGGAGACAGTA 1680
Db 2228 AATGACTTCCAGTGTGATGAATCAATAACTTCAATTTGTGACAGAAAGGAGGCGAGTA 2287
Qy 1681 CTGTCATCTGCAATTA 1695
Db 2288 CCATCATCATATTA 2302

RESULT 8
AB052103 4330 bp mRNA linear PRI 08-MAR-2001
LOCUS Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
DEFINITION type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
JOURNAL
MEDLINE 21092718
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
School of Medicine, Division of Biochemistry, Biomedical Research
Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)
FEATURES
Location/Qualifiers
Source 1. 4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1. 4330
/gene="SRCL"
CDS 3. 1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:13365553"
/translation="MKDDFAEEVPSFGFKRFGHEGTCQPKCHNNALKFSILLY
ILCALLTITVAILGYKVVEKMDNVTGGMETSRQVYDDKLTAVESDLKILGDOTGKKA
STNSELSTFRSDILDRLQQLREITEKTSKNDLLEKLAQSGDALVDROSQKLENN
SFLITTVNKTLOAYVYNLQQDTSLVGLNQNLNOMYSHNVVNNLNLTQVOORN
LITNLRSDVDTQAIRIKNDQVFLQAKKDTDLKLEKVKQSLQTLAANSALA
KANNDLTEDMNSQLNSFTGOMENTTISQANEQNLKLDQLHKAENRTAKFNOLKE
RFLQFTDIVNIISNYSYTAHLRTLSNLNEVRTCTDTLTHTKDDTLTNNLTANI
RLDSVLRMQODLMSRLDTEVANSVIMEEMKLVDSKHGOLIKFTTLOGPGRGP
RGDRGSGPPGTNGKGKBEKBPGRPGPAGERGPIGPAPGRGKSGKSGQPK
GSRGSPGKPGPGGDPGPPGKGLPGPPGPGGLOGTGTGEPVGPGRGLPG
LPVGPMPGPKGPPGPPGPPGAVPLALQNEPTTAPEDNSKSPSLQPGGGSACA"
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN
Query Match 75.6%; Score 1281.2; DB 9; Length 4330;
Best Local Similarity 99.8%; Pred. No. 2.2e-291;
```

Matches 1283; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1	GTCAAGATCTGCAGCAAGATACACGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT	60
Db	534	GTCAAGATCTGCAGCAAGATACACGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT	593
QY	61	TCTCATAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG	120
Db	594	TCTCATAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG	653
QY	121	AACTCATACGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC	180
Db	654	AACTCATACGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC	713
QY	181	AAGAAGCACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGGACCGAATGG	240
Db	714	AAGAAGCACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGGACCGAATGG	773
QY	241	CTGAAGGAGAAAGTGCAGAGCTTTGCAGAGCTGGCTGCCAACAACCTCTGCTTGGCCAAA	300
Db	774	CTGAAGGAGAAAGTGCAGAGCTTTGCAGAGCTGGCTGCCAACAACCTCTGCTTGGCCAAA	833
QY	301	GCCAAACACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAAGTG	360
Db	834	GCCAAACACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAAGTG	893
QY	361	GAGAACATCACTACTATCTCTCAAGCCCAAGCAGCAACCTGAAAGACCTGCAGGACTTA	420
Db	894	GAGAACATCACTACTATCTCTCAAGCCCAAGCAGCAACCTGAAAGACCTGCAGGACTTA	953
QY	421	CACAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG	480
Db	954	CACAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG	1013
QY	481	CTCTTGAGACGGATTTGTGAACATCATTAGCAATATCAGTTACACAGCCACACCTG	540
Db	1014	CTCTTGAGACGGATTTGTGAACATCATTAGCAATATCAGTTACACAGCCACACCTG	1073
QY	541	CGGACGCTGACCAACATCTAAATGAAGTCAGGACCACTTGACAGATACCCCTTACCAAA	600
Db	1074	CGGACGCTGACCAACATCTAAATGAAGTCAGGACCACTTGACAGATACCCCTTACCAAA	1133
QY	601	CACACAGATCTGACCTCTTGAATAATACCTGGCCAACTCCGTTGGATCTGTT	660
Db	1134	CACACAGATCTGACCTCTTGAATAATACCTGGCCAACTCCGTTGGATCTGTT	1193
QY	661	TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAAGTAGCCAACTTA	720
Db	1194	TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAAGTAGCCAACTTA	1253
QY	721	TCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT	780
Db	1254	TCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT	1313
QY	781	TTTACAACTACTCAAGGTCCACCGGGCCCGCAGGGTCCAAAGGTGACAGAGATCCCAG	840
Db	1314	TTTACAACTACTCAAGGTCCACCGGGCCCGCAGGGTCCAAAGGTGACAGAGATCCCAG	1373
QY	841	GGACCCCTGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAAAGGGAGCCTGGACCA	900
Db	1374	GGACCCCTGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAAAGGGAGCCTGGACCA	1433
QY	901	CTTGGCCCTGCGGGTGAGAGAGGCCAAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC	960
Db	1434	CTTGGCCCTGCGGGTGAGAGAGGCCAAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC	1493
QY	961	GGCAAGGATCTAAAGGCTCCAGAGCCCAAGAGCTCCCGTGGTTCCCTGGGAAGCCC	1020
Db	1494	GGCAAGGATCTAAAGGCTCCAGAGCCCAAGAGCTCCCGTGGTTCCCTGGGAAGCCC	1553
QY	1021	GGCCCTCAGGGCCCCAGTGGGAGCCAGGCCCCCGGGCCCAAGAGGAGGAGCTC	1080
Db	1554	GGCCCTCAGGGCCCCAGTGGGAGCCAGGCCCCCGGGCCCAAGAGGAGGAGCTC	1613

QY	1081	CCCGGCCCTCAGGGCCCTCTCTGCTTTCCAGGGACTTTCAGGGCACCGTTGGGAGCCTGGG	1140
Db	1614	CCCGGCCCTCAGGGCCCTCTCTGCTTTCCAGGGACTTTCAGGGCACCGTTGGGAGCCTGGG	1673
QY	1141	GTGCTTGAGACCTCGGGGACTTGCAGGCTTTCCTTGGGTACCAAGGCATGCCAGCCCAAG	1200
Db	1674	GTGCTTGAGACCTCGGGGACTTGCAGGCTTTCCTTGGGTACCAAGGCATGCCAGCCCAAG	1733
QY	1201	GGCCCCCGGGCCCTCTCTGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAG	1260
Db	1734	GGCCCCCGGGCCCTCTCTGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAG	1793
QY	1261	CCAAACCCCGGCACCGGAGGACAATGG	1286
Db	1794	CCAAACCCCGGCACCGGAGGACAATAG	1819

RESULT 9
HSM802985
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HSM802985 1886 bp mRNA linear PRI 20-MAR-2002
Homo sapiens mRNA; cDNA DKFp547G1215 (from clone DKFp547G1215).
AL713657
AL713657.1 GI:19584339
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Koehler, K., Beyer, A., Mewes, H.W., Weill, B. and Wiemann, S.
Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFp547G1215) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
Source
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFp547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbrl). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1886
/gene="DKFp547G1215"
<1..1256
/gene="DKFp547G1215"
/note="similarity to asialoglycoprotein receptor (Mus
musculus"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKFNLEERFQLETFDIVNIISNISTYFAHLRLTSLNLEVRT
TCTDPLTKHTDPLTSLNLTANRLDSVLRMOODLMRSLDTEVANLSVIMEKLV
DSKHQLIKNTFIILOGPGRGPRGDRSQPGPTGKQKGEKGEKGPAGERG
PIGPAGPEERGGKSGQSGPKGSPGPGSGSDPGPPGPPGKGLPGQGP
PFGQLOGTGVGPGVPGPLGPGKPGPGPPGPGPSGAVVPLALQNEPTPA
PDNCGCPHWKNTDKCYFVSVEKEIFEDAKLFCEDKSKSLVFIINTREQOIKOMV
GRESHWIGLTDSERENKWLDTGTPDYKNKAGOPDNWNGHGGPCDCAGLIYAGOW
NDFOCEDVNNFICEKDRETVLSSAL"
1838..1843
polyA_signal

```
/gene="DKF2p547G1215"
1869
polyA_site
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN

Query Match 73.8%; Score 1251.4; DB 9; Length 1886;
Best Local Similarity 99.9%; Pred. No. 2.1e-284;
Matches 1252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 443 CAGCCATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGGATATTGTCA 502
DB 1 CAGCCATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGGATATTGTCA 60
QY 503 ACATCATTAGCAATATCAGTTACACAGCCACCACCTCGCGACGCTTGACCAACAATCTAA 562
DB 61 ACATCATTAGCAATATCAGTTACACAGCCACCACCTCGCGACGCTTGACCAACAATCTAA 120
QY 563 ATGAAGTCAGGACCACTTCCACAGATACCTTTACCAACACACAGATGATCTGACCTCT 622
DB 121 ATGAAGTCAGGACCACTTCCACAGATACCTTTACCAACACACAGATGATCTGACCTCT 180
QY 623 TGAATAATACCTCGGCCAACATCCGTTTTCGATTCTCTCTCAGGATGCAACAGATT 682
DB 181 TGAATAATACCTCGGCCAACATCCGTTTTCGATTCTCTCTCAGGATGCAACAGATT 240
QY 683 TGATGAGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGA 742
DB 241 TGATGAGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGA 300
QY 743 AGCTAGTAGCTCCCAAGCATGTTGCTCATCAAGAAATTTTCAATACATACAGGTTCCAC 802
DB 301 AGCTAGTAGCTCCCAAGCATGTTGCTCATCAAGAAATTTTCAATACATACAGGTTCCAC 360
QY 803 CGGGCCCCAGGGTCCCAAGAGTGACAGAGGATCCAGGAGCCCCCTGCCCAACTGGCA 862
DB 361 CGGGCCCCAGGGTCCCAAGAGTGACAGAGGATCCAGGAGCCCCCTGCCCAACTGGCA 420
QY 863 ACAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACACCTCGCCCTCGGGGTGAGAGAG 922
DB 421 ACAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACACCTCGCCCTCGGGGTGAGAGAG 480
QY 923 GCCCAATTGGACAGCTGTTCCCGCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCC 982
DB 481 GCCCAATTGGACAGCTGTTCCCGCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCC 540
QY 983 AGGGCCCCAAGGCTCCCTGTTCCCTTGGAGAGCCCGCCCTCAGGCCCCCAGTGGGG 1042
DB 541 AGGGCCCCAAGGCTCCCTGTTCCCTTGGAGAGCCCGCCCTCAGGCCCCCAGTGGGG 600
QY 1043 ACCAGGCCCCCGGGCCACCAGGCAAGAGGAGTCCCGGCCCTCAGGCCCCCTCCCTG 1102
DB 601 ACCAGGCCCCCGGGCCACCAGGCAAGAGGAGTCCCGGCCCTCAGGCCCCCTCCCTG 560
QY 1103 GCTTCAGGAGACTTCAGGACACCGTTTGGGAGCCCTGGGTGCTCGGACCTCGGGGACTGC 1162
DB 661 GCTTCAGGAGACTTCAGGACACCGTTTGGGAGCCCTGGGTGCTCGGACCTCGGGGACTGC 720
QY 1163 CAGGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGCCCCCGCCCTCCTGGCC 1222
DB 721 CAGGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGCCCCCGCCCTCCTGGCC 780
QY 1223 CATCAGGAGCGTGGTGGCCCTCGCCCTGCAGAAATGAGCAACCCCGGACCGGAGGACA 1282
DB 781 CATCAGGAGCGTGGTGGCCCTCGCCCTGCAGAAATGAGCAACCCCGGACCGGAGGACA 840
QY 1283 ATGGCTGCCCGCTCAGTGGAAAGACTTCACAGACAAATGCTACTATTTTTCAGTTGAGA 1342
DB 841 ATGGCTGCCCGCTCAGTGGAAAGACTTCACAGACAAATGCTACTATTTTTCAGTTGAGA 900
QY 1343 AGAAATTTTTCAGATGCAAGCTTTTCTGTGAGACAAAGTCTCACATCTGTTTCA 1402
DB 901 AAGAAATTTTTCAGATGCAAGCTTTTCTGTGAGACAAAGTCTCACATCTGTTTCA 960
```

QY 1403 TAAACACTAGAGGAGAACCAATGATGATAAAAAACAGATGTTAGGAGAGAGCCACT 1462
DB 961 TAAACACTAGAGGAGAACCAATGATGATAAAAAACAGATGTTAGGAGAGAGCCACT 1020
QY 1463 GGATGGCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTGGATGGGACATCTC 1522
DB 1021 GGATGGCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTGGATGGGACATCTC 1080
QY 1523 CAGACTACAAAAATTTGGAAGCTGGACACCGGATAACTGGGTCTATGCCATGGGCCAG 1582
DB 1081 CAGACTACAAAAATTTGGAAGCTGGACACCGGATAACTGGGTCTATGCCATGGGCCAG 1140
QY 1583 GAGAAGACTGTCTGGGTTGATTTATGCTGGCAGTGAACGATTTCCTAATGTGAAGAC 1642
DB 1141 GAGAAGACTGTCTGGGTTGATTTATGCTGGCAGTGAACGATTTCCTAATGTGAAGAC 1200
QY 1643 TCAATAACTTTCATTCGGAAGAAAGACAGGACAGACTACTGTCATCTGCATTA 1695
DB 1201 TCAATAACTTTCATTCGGAAGAAAGACAGGACAGACTACTGTCATCTGCATTA 1253

RESULT 10
AC016128/c
LOCUS
DEFINITION
AC016128
VERSION
AC016128.4 GI:10046526
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-324G2
Unpublished
2 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhaltier,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hago,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L999
Center clone name: 324_G2
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165522 bases at least Q40

Consensus quality: 167248 bases at least Q30
 Consensus quality: 167996 bases at least Q20
 Insert size: 177000; agarose-fp
 Insert size: 168388; sum-of-contigs
 Quality coverage: 6.2 in Q20 bases; agarose-fp
 Quality coverage: 6.5 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 17943: contig of 17943 bp in length
 * 17944 18043: gap of 100 bp
 * 18044 22424: contig of 4381 bp in length
 * 22425 22524: gap of 100 bp
 * 22525 38094: contig of 15570 bp in length
 * 38095 38194: gap of 100 bp
 * 38195 49220: contig of 11026 bp in length
 * 49221 49320: gap of 100 bp
 * 49321 66394: contig of 17074 bp in length
 * 66395 66494: gap of 100 bp
 * 66495 91692: contig of 25198 bp in length
 * 91693 91792: gap of 100 bp
 * 91793 120869: contig of 29077 bp in length
 * 120870 120969: gap of 100 bp
 * 120970 169088: contig of 48119 bp in length.

FEATURES

```

source
  1. .169088
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="18"
    /map="18"
    /clone="RP11-324G2"
    /clone_lib="RPC1-11 Human Male BAC"
misc_feature
  1. .17943
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left
misc_feature
  18044..22424
    /note="assembly_fragment"
    22525..38094
    /note="assembly_fragment"
    38195..49220
    /note="assembly_fragment"
    49321..66394
    /note="assembly_fragment"
    66495..91692
    /note="assembly_fragment"
    91793..120869
    /note="assembly_fragment"
    120970..169088
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN
Query Match 47.1%; Score 798; DB 2; Length 169088;
Best Local Similarity 100.0%; Pred. No. 3.8e-177;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCACGAATCTGCAGCAAGATACACGGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db 165666 GTCACGAATCTGCAGCAAGATACACGGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 165607
QY 61 TCTCATTAATGTGTCATCATGAACCTCAACAACTGAACCTGACCCAGGTGCAGCAGAGG 120
Db 165606 TCTCATTAATGTGTCATCATGAACCTCAACAACTGAACCTGACCCAGGTGCAGCAGAGG 165547

```

```

QY 121 AACCTCATCACGAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
Db 165546 AACCTCATCACGAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 165487
QY 181 AGAAGCAGCTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGGACAGCGATTGG 240
Db 165486 AGAAGCAGCTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGGACAGCGATTGG 165427
QY 241 CTGAAGGAGAAAGTCAGAGCTTGCAGAGCTGGCTGCCAAACAACTCTGCGTTGGGCCAAA 300
Db 165426 CTGAAGGAGAAAGTCAGAGCTTGCAGAGCTGGCTGCCAAACAACTCTGCGTTGGGCCAAA 165367
QY 301 GCCAACACGACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAAGT 360
Db 165366 GCCAACACGACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAAGT 165307
QY 361 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCGAGGACTTA 420
Db 165306 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCGAGGACTTA 165247
QY 421 CACAAAGATCGAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG 480
Db 165246 CACAAAGATCGAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG 165187
QY 481 CTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTTACAGAGCCCAACCTG 540
Db 165186 CTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTTACAGAGCCCAACCTG 165127
QY 541 CGGACCTGACCAAGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCCCTTACCAA 600
Db 165126 CGGACCTGACCAAGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCCCTTACCAA 165067
QY 601 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCGGTTTGGATTCTGTT 660
Db 165066 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCGGTTTGGATTCTGTT 165007
QY 661 TCTCTCAGGATCAACAAGATTTGATGAGTCTGAGGTGTAGACACTGAAGTAGCAACTTA 720
Db 165006 TCTCTCAGGATCAACAAGATTTGATGAGTCTGAGGTGTAGACACTGAAGTAGCAACTTA 164947
QY 721 TCAGTGATTATGAAGAAATGAAGCTAGTAGCTCAAGCATGTCAGCTCATCAAGAAT 780
Db 164946 TCAGTGATTATGAAGAAATGAAGCTAGTAGCTCAAGCATGTCAGCTCATCAAGAAT 164887
QY 781 TTTACAATACTACAAGGT 798
Db 164886 TTTACAATACTACAAGGT 164869

RESULT 11
AP000900 178022 bp DNA linear HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKING
DEFINITION DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION AP000900
VERSION AP000900.3 GI:8119043
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-683J11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 178022)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 178,022 genomic DNA of 18p11.3
PUBLISHED Only in DataBase (1999)
REFERENCE 2 (bases 1 to 178022)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

```

Japan (E-mail: hattori@gsr.riken.go.jp,
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924)

On May 31, 2000 this sequence version replaced gi:6997403.

COMMENT

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsr.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RPL1-683J11

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 161944 bases at least Q40

Consensus quality: 170357 bases at least Q30

Consensus quality: 174322 bases at least Q20

Insert size: 176122; sum-of-contigs

Quality coverage: 4.56x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 24283 contig of 24283 bp in length
24384 46875 contig of 22492 bp in length
46976 65058 contig of 18083 bp in length
65159 81322 contig of 16164 bp in length
81423 98823 contig of 17401 bp in length
98924 110068 contig of 11145 bp in length
110169 122716 contig of 12548 bp in length
122817 133089 contig of 10273 bp in length
133190 141206 contig of 8017 bp in length
141307 149050 contig of 7744 bp in length
149151 156105 contig of 6955 bp in length
156206 162048 contig of 5843 bp in length
162149 165118 contig of 2970 bp in length
165219 167480 contig of 2262 bp in length
167581 169986 contig of 2406 bp in length
169987 171909 contig of 1823 bp in length
172010 172317 contig of 308 bp in length
172418 174663 contig of 2246 bp in length
174764 176271 contig of 1508 bp in length
176372 178022 contig of 1651 bp in length

```

Sequence updated (24-Dec-1999)

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 24283: contig of 24283 bp in length
* 24284 24383: gap of 100 bp
* 24384 46875: contig of 22492 bp in length
* 46876 46975: gap of 100 bp
* 46976 65058: contig of 18083 bp in length
* 65059 65158: gap of 100 bp
* 65159 81322: contig of 16164 bp in length
* 81323 81422: gap of 100 bp
* 81423 98823: contig of 17401 bp in length
* 98824 98923: gap of 100 bp
* 98924 110068: contig of 11145 bp in length
* 110069 110168: gap of 100 bp
* 110169 122716: contig of 12548 bp in length
* 122717 122816: gap of 100 bp

```

```

* 122817 133089: contig of 10273 bp in length
* 133090 133189: gap of 100 bp
* 133190 141206: contig of 8017 bp in length
* 141207 141306: gap of 100 bp
* 141307 149050: contig of 7744 bp in length
* 149051 149150: gap of 100 bp
* 149151 156105: contig of 6955 bp in length
* 156106 156205: gap of 100 bp
* 156206 162048: contig of 5843 bp in length
* 162049 162148: gap of 100 bp
* 162149 165118: contig of 2970 bp in length
* 165119 165218: gap of 100 bp
* 165219 167480: contig of 2262 bp in length
* 167481 167580: gap of 100 bp
* 167581 169986: contig of 2406 bp in length
* 169987 170086: gap of 100 bp
* 170087 171909: contig of 1823 bp in length
* 171910 172009: gap of 100 bp
* 172010 172317: contig of 308 bp in length
* 172318 172417: gap of 100 bp
* 172418 174663: contig of 2246 bp in length
* 174664 174763: gap of 100 bp
* 174764 176271: contig of 1508 bp in length
* 176272 176371: gap of 100 bp
* 176372 178022: contig of 1651 bp in length.

```

FEATURES

Source

1. 178022

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18p11.3"

/clone="RPL1-683J11"

1. 24283

misc_feature /note="assembly_fragment"

misc_feature 24384..46875

misc_feature /note="assembly_fragment"

misc_feature 46976..65058

misc_feature /note="assembly_fragment"

misc_feature 65159..81322

misc_feature /note="assembly_fragment"

misc_feature 81423..98823

misc_feature /note="assembly_fragment"

misc_feature 98924..110068

misc_feature /note="assembly_fragment"

misc_feature 110169..122716

misc_feature /note="assembly_fragment"

misc_feature 122817..133089

misc_feature /note="assembly_fragment"

misc_feature 133190..141206

misc_feature /note="assembly_fragment"

misc_feature 141307..149050

misc_feature /note="assembly_fragment"

misc_feature 149151..156105

misc_feature /note="assembly_fragment"

misc_feature 156206..162048

misc_feature /note="assembly_fragment"

misc_feature 162149..165118

misc_feature /note="assembly_fragment clone_end:SP6 vector_side:left"

misc_feature 165219..167480

misc_feature /note="assembly_fragment"

misc_feature 167581..169986

misc_feature /note="assembly_fragment"

misc_feature 170087..171909

misc_feature /note="assembly_fragment"

misc_feature 172010..172317

misc_feature /note="assembly_fragment clone_end:T7 vector_side:right"

misc_feature 172418..174663

misc_feature /note="assembly_fragment"

misc_feature 174764..176271

misc_feature /note="assembly_fragment"

misc_feature 176372..178022

misc_feature /note="assembly_fragment"

BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others

ORIGIN

Query Match 47.1%; Score 798; DB 2; Length 178022;
Best Local Similarity 100.0%; Pred. No. 3.8e-177;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GTCAAGNAATCTGCAGCAAGATACACGGTCTCCAGGCGAATCTGCAGAACCAAAATGTAT	60
Db	13104	GTCAAGNAATCTGCAGCAAGATACACGGTCTCCAGGCGAATCTGCAGAACCAAAATGTAT	13163
QY	61	TCTCATTAATCTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG	120
Db	13164	TCTCATTAATCTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG	13223
QY	121	AACTCATACAGAAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC	180
Db	13224	AACTCATACAGAAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC	13283
QY	181	AAGAACGACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACCGATTGG	240
Db	13284	AAGAACGACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACCGATTGG	13343
QY	241	CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAACAACTCTGGTTGGCCAAA	300
Db	13344	CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAACAACTCTGGTTGGCCAAA	13403
QY	301	GCCAAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG	360
Db	13404	GCCAAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG	13463
QY	361	GAGAAATCACCACCTATCTCAAGCCAGGACGACGAACCTGAAAGACCTGCAGGACTTA	420
Db	13464	GAGAAATCACCACCTATCTCAAGCCAGGACGACGAACCTGAAAGACCTGCAGGACTTA	13523
QY	421	CACAAAGATCGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG	480
Db	13524	CACAAAGATCGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG	13583
QY	481	CTCTTTGAGACGGATATTGTGAACATCAATTAGCAATATCAGTTTACAGCCCCACACCTG	540
Db	13584	CTCTTTGAGACGGATATTGTGAACATCAATTAGCAATATCAGTTTACAGCCCCACACCTG	13643
QY	541	CGGACCTGACCAAGCAATCTAAATGATGAGTCAGGACCACTTGACACATACCCCTTACCAA	600
Db	13644	CGGACCTGACCAAGCAATCTAAATGATGAGTCAGGACCACTTGACACATACCCCTTACCAA	13703
QY	601	CACACAGATGATCTGACCTCTTGAATAATACCTTGCCCAACATCCGTTTGGATTCTGTT	660
Db	13704	CACACAGATGATCTGACCTCTTGAATAATACCTTGCCCAACATCCGTTTGGATTCTGTT	13763
QY	661	TCTCTCAGGATGCAACAAGATTTGATGAGGTGAGGTTAGACACTGAAGTAGCAACTTA	720
Db	13764	TCTCTCAGGATGCAACAAGATTTGATGAGGTGAGGTTAGACACTGAAGTAGCAACTTA	13823
QY	721	TCAGTGATTTGAAAGAAATGAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT	780
Db	13824	TCAGTGATTTGAAAGAAATGAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT	13883
QY	781	TTTACAAATCTACAAGGT 798	
Db	13884	TTTACAAATCTACAAGGT 13901	

RESULT 12
AP001022
LOCUS AP001022 187635 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING
DRAFT SEQUENCE, 33 unordered pieces.
ACCESSION AP001022
VERSION AP001022.2 GI:8117692
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-815L4.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 187,635 genomic DNA of 18p11.3
Published Only in DataBase (2000)
2 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997772.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: Humdraft18
Center clone name: RP11-815L4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167192 bases at least Q40
Consensus quality: 177020 bases at least Q30
Consensus quality: 182105 bases at least Q20
Insert size: 184435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
33 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1	21332	contig of	21332	bp in	length
21433	42072	contig of	20640	bp in	length
42173	54336	contig of	12164	bp in	length
54437	66766	contig of	12330	bp in	length
66867	77849	contig of	10983	bp in	length
77950	87798	contig of	9849	bp in	length
87899	99510	contig of	11612	bp in	length
99611	107273	contig of	7663	bp in	length
107374	113189	contig of	5816	bp in	length
113290	120559	contig of	7270	bp in	length
120660	126468	contig of	5809	bp in	length
128569	131619	contig of	5051	bp in	length
131720	136810	contig of	5091	bp in	length
136911	139925	contig of	3015	bp in	length
140026	144232	contig of	4207	bp in	length
144333	148902	contig of	4570	bp in	length
149003	152918	contig of	3916	bp in	length
153019	156399	contig of	3381	bp in	length
156500	158933	contig of	2434	bp in	length
159034	161884	contig of	2851	bp in	length
161985	164301	contig of	2317	bp in	length
164402	166300	contig of	1899	bp in	length
166401	168580	contig of	2180	bp in	length
168681	171270	contig of	2590	bp in	length
171371	174092	contig of	2722	bp in	length
174193	176095	contig of	1903	bp in	length
176196	177890	contig of	1695	bp in	length
177991	179786	contig of	1796	bp in	length
179887	181478	contig of	1592	bp in	length
181579	182895	contig of	1317	bp in	length

182996 184841 contig of 1846 bp in length
184942 186345 contig of 1404 bp in length
186446 187635 contig of 1190 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```
1 21332: contig of 21332 bp in length
* 21333 21432: gap of 100 bp
* 21433 42072: contig of 20640 bp in length
* 42073 42172: gap of 100 bp
* 42173 54336: contig of 12164 bp in length
* 54337 54436: gap of 100 bp
* 54437 66766: contig of 12330 bp in length
* 66767 68666: gap of 100 bp
* 68667 77849: contig of 10983 bp in length
* 77850 77949: gap of 100 bp
* 77950 87798: contig of 9849 bp in length
* 87799 87898: gap of 100 bp
* 87899 99510: contig of 11612 bp in length
* 99511 99610: gap of 100 bp
* 99611 107273: contig of 7663 bp in length
* 107274 107373: gap of 100 bp
* 107374 113189: contig of 5816 bp in length
* 113190 113289: gap of 100 bp
* 113290 120559: contig of 7270 bp in length
* 120600 120659: gap of 100 bp
* 120660 126468: contig of 5809 bp in length
* 126469 126568: gap of 100 bp
* 126569 131619: contig of 5051 bp in length
* 131620 131719: gap of 100 bp
* 131720 136810: contig of 5091 bp in length
* 136811 136910: gap of 100 bp
* 136911 139925: contig of 3015 bp in length
* 139926 140025: gap of 100 bp
* 140026 144232: contig of 4207 bp in length
* 144233 144332: gap of 100 bp
* 144333 148902: contig of 4570 bp in length
* 148903 149002: gap of 100 bp
* 149003 152918: contig of 3916 bp in length
* 152919 153018: gap of 100 bp
* 153019 156399: contig of 3381 bp in length
* 156400 156499: gap of 100 bp
* 156500 158933: contig of 2434 bp in length
* 158934 159033: gap of 100 bp
* 159034 161884: contig of 2851 bp in length
* 161885 161984: gap of 100 bp
* 161985 164301: contig of 2317 bp in length
* 164302 164401: gap of 100 bp
* 164402 166300: contig of 1899 bp in length
* 166301 166400: gap of 100 bp
* 166401 168580: contig of 2180 bp in length
* 168581 168680: gap of 100 bp
* 168681 171270: contig of 2590 bp in length
* 171271 171370: gap of 100 bp
* 171371 174092: contig of 2722 bp in length
* 174093 174192: gap of 100 bp
* 174193 176095: contig of 1903 bp in length
* 176096 176195: gap of 100 bp
* 176196 177890: contig of 1695 bp in length
* 177891 177990: gap of 100 bp
* 177991 179786: contig of 1796 bp in length
* 179787 179886: gap of 100 bp
* 179887 181478: contig of 1592 bp in length
* 181479 181578: gap of 100 bp
* 181579 182895: contig of 1317 bp in length
* 182896 182995: gap of 100 bp
* 182996 184841: contig of 1846 bp in length
```

```

* 184842 184941: gap of 100 bp
* 184942 186345: contig of 1404 bp in length
* 186346 186445: gap of 100 bp
* 186446 187635: contig of 1190 bp in length.

FEATURES             source
1..187635             Location/Qualifiers
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="18"
                        /map="18p11.3"
                        /clone="RP11-815L4"
1..21332             misc_feature
21433..42072         misc_feature
42173..54336         misc_feature
54437..66766         misc_feature
66867..77849         misc_feature
77950..87798         misc_feature
87899..99510         misc_feature
99611..107273        misc_feature
107374..113189       misc_feature
113290..120559       misc_feature
120660..126468       misc_feature
126569..131619       misc_feature
131720..136810       misc_feature
136911..139925       misc_feature
140026..144232       misc_feature
144333..148902       misc_feature
149003..152918       misc_feature
153019..156399       misc_feature

Query Match          47.1%; Score 798; DB 2; Length 187635;
Best Local Similarity 100.0%; Pred. No. 3.8e-177;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  GTCACGAATCTGCAGCAAGATACCGCGTGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db 58354 GTCACGAATCTGCAGCAAGATACCGCGTGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 58413

QY 61  TCTCATAATGTGGTTCATCATGAACCTCAACAACTGAACCTGACCCAGGTGCAGCAGAG 120
Db 58414 TCTCATAATGTGGTTCATCATGAACCTCAACAACTGAACCTGACCCAGGTGCAGCAGAG 58473

QY 121 AACCTCATCAGNAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC 180
Db 58474 AACCTCATCAGNAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC 58533

QY 181 AGAAGCACTTTCAAATCTGCAGCAGGTTTTTTCTTCAAGCCAAAGAGGACACGGATTGG 240
Db 58534 AGAAGCACTTTCAAATCTGCAGCAGGTTTTTTCTTCAAGCCAAAGAGGACACGGATTGG 58593

QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTCGCTTGGCCAAA 300
Db 58594 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTCGCTTGGCCAAA 58653

QY 301 GCCAACACGACACCTCGGAGGATATGACAGCCAGCTCAACTCACTTACAGGTCAGATG 360
|||||
```

Db	58654	CCCAACAGCAGCACCCCTGGAGGATATGAACAGCAGCTCAACTCATTCACAGGTGAGATG	58713
Qy	361	GAGAACATCACCACTATCTCTCAAGCCAAACGAGCAGAACCTGAAAGACCTGCAGGACTTA	420
Db	58714	GAGAACATCACCACTATCTCTCAAGCCAAACGAGCAGAACCTGAAAGACCTGCAGGACTTA	58773
Qy	421	CACAAGATGCAGAGATATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG	480
Db	58774	CACAAGATGCAGAGATATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG	58833
Qy	481	CTCTTTGAGCGGATATGTGAACATCATTAGCAATATCAGTTACACAGCCCAACCTG	540
Db	58834	CTCTTTGAGCGGATATGTGAACATCATTAGCAATATCAGTTACACAGCCCAACCTG	58893
Qy	541	CGGACGCTGACCAACAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA	600
Db	58894	CGGACGCTGACCAACAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA	58953
Qy	601	CACACAGATGATCGACCTCCTTGAATAATAACCTGCCAACATCCGTTTGATTCCTGTT	660
Db	58954	CACACAGATGATCGACCTCCTTGAATAATAACCTGCCAACATCCGTTTGATTCCTGTT	59013
Qy	661	TCTCTCAGGATGCAACAGATTTCATGAGTCGAGGTTAGACATGAAGTAGGCAACTTA	720
Db	59014	TCTCTCAGGATGCAACAGATTTCATGAGTCGAGGTTAGACATGAAGTAGGCAACTTA	59073
Qy	721	TCAGTGTATTGGAGAAATCAAGCTAGTACATCCCAAGCATGCTCAGCTCATCAAGAAT	780
Db	59074	TCAGTGTATTGGAGAAATCAAGCTAGTACATCCCAAGCATGCTCAGCTCATCAAGAAT	59133
Qy	781	TTTACAATACTACAAGGT 798	
Db	59134	TTTACAATACTACAAGGT 59151	
RESULT 13	AP000915/c		
LOCUS			
DEFINITION	Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete sequences.		
ACCESSION	AP000915		
VERSION	AP000915.5	GI:20334314	
KEYWORDS	HTG.		
SOURCE			
ORGANISM	Homo sapiens DNA, clone:RP11-720L2.		
REFERENCE			
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Homo sapiens genomic DNA		
JOURNAL	Published Only in Database (1999)		
REFERENCE	2 (bases 1 to 188439)		
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-DEC-1999), Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
COMMENT	On Apr 26, 2002 this sequence version replaced gi:9188470.		
FEATURES	Location/Qualifiers		
source	1..188439		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="18"		
	/map="18p"		
	/clone="RP11-720L2"		
BASE COUNT	54003 a 41134 c 40027 g 53275 t		
ORIGIN			
Query Match	47.1%	Score 798;	DB 9; Length 188439;

Best Local Similarity 100.0%; Pred. No. 3.8e-177; Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GTCAGGAATCTGCAGCAAGATACCAAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db	56736 GTCAGGAATCTGCAGCAAGATACCAAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 56737
QY	61 TCTCATAATGGTTCATCATGAACCTCAACACCTGAACCTGACCCAGTGCAGCAGAG 120
Db	56736 TCTCATAATGGTTCATCATGAACCTCAACACCTGAACCTGACCCAGTGCAGCAGAG 56737
QY	121 AACCTCATCAAGAATCTCAGCGGTCTGTGATGACACAAAGCCAGGCTATCCAGCGAATC 180
Db	56676 AACCTCATCAAGAATCTCAGCGGTCTGTGATGACACAAAGCCAGGCTATCCAGCGAATC 56677
QY	181 AAGACGACTTTCAAATCTGCAGAGGTTTTTCTTCAAGCCAAAGAGCACGGATTGG 240
Db	56616 AAGACGACTTTCAAATCTGCAGAGGTTTTTCTTCAAGCCAAAGAGCACGGATTGG 56557
QY	241 CTGAAGGAGAAGTGCAGAGCTTGCAGAGCTGGCTGCGCAAACTCTGCGTTGGCCAAA 300
Db	56556 CTGAAGGAGAAGTGCAGAGCTTGCAGAGCTGGCTGCGCAAACTCTGCGTTGGCCAAA 56497
QY	301 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG 360
Db	56496 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG 56437
QY	361 GAGAACATCACCACTATCTCTCAAGCCAAACAGCAGAACCTGAAGAGCTTCAGGACTTA 420
Db	56436 GAGAACATCACCACTATCTCTCAAGCCAAACAGCAGAACCTGAAGAGCTTCAGGACTTA 56377
QY	421 CACAAGATGCAGAGAATAGAACACGCCATCAAGTTCAACCAACTGGAGGAACGCTTCAG 480
Db	56376 CACAAGATGCAGAGAATAGAACACGCCATCAAGTTCAACCAACTGGAGGAACGCTTCAG 56317
QY	481 CTCTTTGAGCGGATATTTGTGAACATCATTAGCAATATCAGTTACACAGCCCAACCTG 540
Db	56316 CTCTTTGAGCGGATATTTGTGAACATCATTAGCAATATCAGTTACACAGCCCAACCTG 56257
QY	541 CGGAGCTGCACCAACTCTAAATGAAGTCAGGACCCTTGCACAGATACCCTTACCAA 600
Db	56256 CGGAGCTGCACCAACTCTAAATGAAGTCAGGACCCTTGCACAGATACCCTTACCAA 56197
QY	601 CACACAGATGATCTGACCTCTTGAATAATACCTTGGCCAAACATCCGTTTGGATTCTGTT 660
Db	56196 CACACAGATGATCTGACCTCTTGAATAATACCTTGGCCAAACATCCGTTTGGATTCTGTT 56137
QY	661 TCTCTCAGGATGCAACAGATTTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTTA 720
Db	56136 TCTCTCAGGATGCAACAGATTTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTTA 56077
QY	721 TCAGTGATTATGGAGAAATGAAGCTAGTACCTCCAGCATGCTCAGCTCATCAAGAT 780
Db	56076 TCAGTGATTATGGAGAAATGAAGCTAGTACCTCCAGCATGCTCAGCTCATCAAGAT 56017
QY	781 TTTACAATACTACAAGGT 798
Db	56016 TTTACAATACTACAAGGT 55999
.	
RESULT	14
LOCUS	AP000939
DEFINITION	Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION	AP000939
VERSION	AP000939.3 GI:8119080
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE	Homo sapiens DNA, clone:RP11-839023.
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 188255)

AUTHORS

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 188,255 genomic DNA of 18p11.3
 Published Only in Database (1999)
 2 (bases 1 to 188255)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail: hattori@gsc.riken.go.jp,
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:6997751.

COMMENT

----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft18
 Center clone name: RP11-839023
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 171068 bases at least Q40
 Consensus quality: 178865 bases at least Q30
 Consensus quality: 182445 bases at least Q20
 Insert size: 185155; sum-of-contigs
 Quality coverage: 4.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 16096 contig of 16096 bp in length
 16197 33789 contig of 17593 bp in length
 33890 51010 contig of 17121 bp in length
 51111 63558 contig of 12448 bp in length
 63659 75056 contig of 11398 bp in length
 75157 84737 contig of 9581 bp in length
 84838 90909 contig of 6072 bp in length
 91010 98615 contig of 7606 bp in length
 98716 105978 contig of 6971 bp in length
 106079 113049 contig of 6052 bp in length
 113150 119201 contig of 7281 bp in length
 119302 126582 contig of 5388 bp in length
 126683 132070 contig of 4968 bp in length
 132171 137138 contig of 5817 bp in length
 137239 143055 contig of 5484 bp in length
 143156 148639 contig of 4153 bp in length
 148740 153086 contig of 3792 bp in length
 153187 158921 contig of 2954 bp in length
 158922 159021 contig of 2210 bp in length
 159022 163174 contig of 100 bp in length
 163175 163274 contig of 100 bp in length
 163275 167066 contig of 392 bp in length
 167067 167166 contig of 100 bp in length
 167167 170120 contig of 2954 bp in length
 170121 170220 contig of 100 bp in length
 170221 172430 contig of 2210 bp in length
 172431 172530 contig of 100 bp in length
 172531 175598 contig of 3068 bp in length
 175599 175698 contig of 100 bp in length
 175699 177118 contig of 1420 bp in length
 177119 177218 contig of 100 bp in length
 177219 178531 contig of 1313 bp in length
 178532 178631 contig of 100 bp in length
 178632 179857 contig of 1226 bp in length
 179858 179957 contig of 100 bp in length
 179958 181064 contig of 1107 bp in length
 181065 181164 contig of 100 bp in length
 181165 182735 contig of 1571 bp in length
 182736 182835 contig of 100 bp in length
 182836 183923 contig of 1088 bp in length
 183924 184023 contig of 100 bp in length
 184024 185810 contig of 1787 bp in length
 185811 185910 contig of 100 bp in length
 185911 187145 contig of 1235 bp in length
 187146 187245 contig of 100 bp in length
 187246 188255 contig of 1010 bp in length

Sequence updated (06-Jan-2000)

FEATURES

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 16096: contig of 16096 bp in length
 * 16097 16196: gap of 100 bp
 * 16197 33789: contig of 17593 bp in length
 * 33790 33889: gap of 100 bp
 * 33890 51010: contig of 17121 bp in length
 * 51011 51110: gap of 100 bp
 * 51111 63558: contig of 12448 bp in length
 * 63559 63658: gap of 100 bp
 * 63659 75056: contig of 11398 bp in length
 * 75057 75156: gap of 100 bp
 * 75157 84737: contig of 9581 bp in length
 * 84738 84837: gap of 100 bp
 * 84838 90909: contig of 6072 bp in length
 * 90910 91009: gap of 100 bp
 * 91010 98615: contig of 7606 bp in length
 * 98616 98715: gap of 100 bp
 * 98716 105978: contig of 7263 bp in length
 * 105979 106078: gap of 100 bp
 * 106079 113049: contig of 6971 bp in length
 * 113050 113149: gap of 100 bp
 * 113150 119201: contig of 6052 bp in length
 * 119202 119301: gap of 100 bp
 * 119302 126582: contig of 7281 bp in length
 * 126583 126682: gap of 100 bp
 * 126683 132070: contig of 5388 bp in length
 * 132071 132170: gap of 100 bp
 * 132171 137138: contig of 4968 bp in length
 * 137139 137238: gap of 100 bp
 * 137239 143055: contig of 5817 bp in length
 * 143056 143155: gap of 100 bp
 * 143156 148639: contig of 5484 bp in length
 * 148640 148739: gap of 100 bp
 * 148740 153086: contig of 4347 bp in length
 * 153087 153186: gap of 100 bp
 * 153187 158921: contig of 5735 bp in length
 * 158922 159021: gap of 100 bp
 * 159022 163174: contig of 4153 bp in length
 * 163175 163274: gap of 100 bp
 * 163275 167066: contig of 392 bp in length
 * 167067 167166: gap of 100 bp
 * 167167 170120: contig of 2954 bp in length
 * 170121 170220: gap of 100 bp
 * 170221 172430: contig of 2210 bp in length
 * 172431 172530: gap of 100 bp
 * 172531 175598: contig of 3068 bp in length
 * 175599 175698: gap of 100 bp
 * 175699 177118: contig of 1420 bp in length
 * 177119 177218: gap of 100 bp
 * 177219 178531: contig of 1313 bp in length
 * 178532 178631: gap of 100 bp
 * 178632 179857: contig of 1226 bp in length
 * 179858 179957: gap of 100 bp
 * 179958 181064: contig of 1107 bp in length
 * 181065 181164: gap of 100 bp
 * 181165 182735: contig of 1571 bp in length
 * 182736 182835: gap of 100 bp
 * 182836 183923: contig of 1088 bp in length
 * 183924 184023: gap of 100 bp
 * 184024 185810: contig of 1787 bp in length
 * 185811 185910: gap of 100 bp
 * 185911 187145: contig of 1235 bp in length
 * 187146 187245: gap of 100 bp
 * 187246 188255: contig of 1010 bp in length.

Location/Qualifiers

REFERENCE
AUTHORS

Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182029)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicoll, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, C., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission

TITLE
JOURNAL

Submitted (10-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 10, 2002 this sequence version replaced gi:21328559.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: U24628
Center clone name: 213_K_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179506 bases at least Q40
Consensus quality: 180723 bases at least Q30
Consensus quality: 181050 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 181129; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-fp
Quality coverage: 8.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 951: contig of 951 bp in length
* 952 1051: gap of 100 bp
* 1052 3806: contig of 2755 bp in length
* 3807 3906: gap of 100 bp
* 3907 8709: contig of 4803 bp in length
* 8710 8809: gap of 100 bp
* 8810 17263: contig of 8454 bp in length
* 17264 17363: gap of 100 bp
* 17364 36938: contig of 19575 bp in length
* 36939 37038: gap of 100 bp
* 37039 53226: contig of 16188 bp in length
* 53227 53326: gap of 100 bp
* 53327 69572: contig of 16246 bp in length

* 69573 69672: gap of 100 bp
* 69673 89021: contig of 19349 bp in length
* 89022 89121: gap of 100 bp
* 89122 128310: contig of 39189 bp in length
* 128311 128410: gap of 100 bp
* 128411 182029: contig of 53619 bp in length.

FEATURES
Location/Qualifiers
Source
1..182029
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-213K19"
/cclone_lib="RPC1-24 Male Mouse BAC"
1..951
/note="assembly_fragment"
1052..3806
/note="assembly_fragment"
3907..8709
/note="assembly_fragment"
8810..17263
/note="assembly_fragment"
17364..36938
/note="assembly_fragment"
37039..53226
/note="assembly_fragment"
53327..69572
/note="assembly_fragment"
69673..89021
/note="assembly_fragment"
89122..128310
/note="assembly_fragment"
128411..182029
/note="assembly_fragment"

BASE COUNT 54245 a 37484 c 36555 g 52844 t 901 others

ORIGIN

Query Match 39.1%; Score 662; DB 2; Length 182029;
Best Local Similarity 89.3%; Pred. No. 4.7e-145;
Matches 713; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1 GTCAGCAATCTGCAGCAAGATACCAGCGTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1931 GTCACAAATCTGCAACAGATAGTACTAGTGTCTCCAGGGCAATCTGCAGAGCCAAATGTAT 1872

QY 61 TCTCAATATGTGTTCATCATGAACCTCAACAACTCAACCTGACCCAGGTGCAGCAGAGG 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1871 TCTCAGAGCGTGTATCATGAACCTCAACAACTCAACCTGACCCAGGTTCAGCAGAGG 1812

QY 121 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1811 AACCTTATCTCAATCTGCAGCAGTCTGTGGATGACACAGCCAGGCTGCGCCATCCAGCGAAT 1752

QY 181 AAGAAGGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGACACGGATTGG 240
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1751 AGAATGATTTCCAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGACACGGATTGG 1692

QY 241 CTCAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGTGCACCAACAACTCTCGCTTGGCCAAA 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1691 CTAAGGAAAAGTACAGAGCTTGCAGACATTTGGTGTGCCAACAACTCTCGCTTGGCCAAA 1632

QY 301 GCCACACGACACCTTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGTTCAGATG 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1631 GCCAACAAATGACACCTTAGAGGATATGAATAGCCAGCTCAGCTCAATTCACAGTTCAGATG 1572

QY 361 GAGACATCCACACTATCTCTCAAGCCAAAGAGCAGCAACTCTGAAGACCTTGCAGGACTTA 420
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1571 GACAACTTACCACCTATCTCAGAGGCAAGCAGCAGAGGCTGAAAGACCTTCAGGACTTA 1512

QY 421 CACAAAGATGACAGAAATAGAACAGCCATCAAGTTCAACCAACTGCAGGAAAGCTTCCAG 480
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1511 CACAGGATACAGAAATAGACAGCTGCAAGTTCAAGCCAACTTCAGGAAAGCTTCCAG 1452

QY 481 CTCCTTTGAGACGGATATCTGCAACATCATTAGCAATATAGTTACAGCCACCACTG 540
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

Db 1451 GTCTTTGAGACAGATATTGTGAACATCATAGCAACATCAGCTACAGCCCATCACCTG 1392
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCCCTTACCAAA 600
Db 1391 AGGACACTGACCAAGCAATCTGAATGATGTTAGGACCACTGCACACACACCTTGACCAGA 1332
QY 601 CACACAGATGATCTGACCTCCTTTGAATAATACCCCTGGCCAAACATCCGTTTGGATTCTGTT 660
Db 1331 CACACGGATGACCTGACCTCCTTTGAATAACACACTAGTCAACATCCGCTTGGATTCTATT 1272
QY 661 TCTCTCAGGATGCAACAAGATTGTGATGAGTGGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTA 720
Db 1271 TCTCTCAGGATGCAACAAGATTGTGATGAGTGGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTA 1212
QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAT 780
Db 1211 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAT 1152
QY 781 TTTACAATACTACAAGGT 798
Db 1151 TTTACCATTCTACAAGGT 1134

```

Search completed: March 21, 2003, 06:24:31
Job time : 5262.67 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 258.28 Seconds
(without alignments)
5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_1_1695

Perfect score: 1695

Sequence: 1 gtcacgaatgtcagcaaga.....cagtactgtcatctgcatta 1695

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*

10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*

12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*

13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1681.4	99.2	2930	10	US-09-745-763-198		Sequence 198, App
2	111	6.5	6728	10	US-09-954-456-782		Sequence 782, App
3	111	6.5	6728	10	US-09-880-107-3946		Sequence 3946, Ap
4	110.6	6.5	1133	9	US-09-924-340-57		Sequence 57, Appl
5	110.6	6.5	1133	9	US-09-992-600A-57		Sequence 57, Appl
6	109.4	6.5	2691	10	US-09-925-302-64		Sequence 64, Appl
7	107.2	6.3	2192	10	US-09-925-301-42		Sequence 42, Appl
8	107	6.3	5086	10	US-09-880-107-3947		Sequence 3947, Ap
9	107	6.3	5145	9	US-09-925-299-206		Sequence 206, App
10	107	6.3	5145	10	US-09-925-299-206		Sequence 206, App
11	107	6.3	5432	12	US-10-044-090-22		Sequence 22, Appl
12	106.8	6.3	4908	9	US-10-001-887-33		Sequence 33, Appl
13	106	6.3	2542	9	US-09-954-531-961		Sequence 961, App
14	106	6.3	2542	10	US-09-964-824A-255		Sequence 255, App
15	105.4	6.2	5416	10	US-09-954-456-786		Sequence 786, App
16	105.4	6.2	5416	10	US-09-880-107-2094		Sequence 2094, Ap
17	105.2	6.2	6158	10	US-09-919-497-6		Sequence 6, Appl
18	105.2	6.2	6158	10	US-09-954-456-762		Sequence 762, App
19	104.2	6.1	4149	12	US-10-044-090-104		Sequence 104, App

20	102.4	6.0	1486	10	US-09-925-302-247	Sequence 247, App
21	101.6	6.0	3226	10	US-09-954-456-725	Sequence 725, App
22	101.4	6.0	3690	12	US-10-044-090-448	Sequence 448, App
23	98.6	5.8	2158	9	US-10-001-835-98	Sequence 98, Appl
24	97.8	5.8	2520	10	US-09-880-107-3685	Sequence 3685, Ap
25	95.8	5.7	1485	10	US-09-925-302-246	Sequence 246, App
26	95.2	5.6	1797	9	US-09-978-295A-613	Sequence 613, App
27	95.2	5.6	1797	9	US-09-978-697-613	Sequence 613, App
28	95.2	5.6	1797	9	US-09-978-192A-613	Sequence 613, App
29	95.2	5.6	1797	9	US-09-999-832A-613	Sequence 613, App
30	95.2	5.6	1797	9	US-09-978-189-613	Sequence 613, App
31	95.2	5.6	1797	9	US-10-174-590-331	Sequence 331, App
32	95.2	5.6	1797	9	US-10-176-758-331	Sequence 331, App
33	95.2	5.6	1797	9	US-10-175-737-331	Sequence 331, App
34	95.2	5.6	1797	9	US-10-173-706-331	Sequence 331, App
35	95.2	5.6	1797	9	US-10-175-738-331	Sequence 331, App
36	95.2	5.6	1797	9	US-10-175-752-331	Sequence 331, App
37	95.2	5.6	1797	9	US-10-176-482-331	Sequence 331, App
38	95.2	5.6	1797	9	US-10-176-757-331	Sequence 331, App
39	95.2	5.6	1797	9	US-10-176-913-331	Sequence 331, App
40	95.2	5.6	1797	9	US-10-180-552-331	Sequence 331, App
41	95.2	5.6	1797	9	US-10-180-557-331	Sequence 331, App
42	95.2	5.6	1797	9	US-10-173-700-331	Sequence 331, App
43	95.2	5.6	1797	9	US-10-174-572-331	Sequence 331, App
44	95.2	5.6	1797	9	US-10-174-579-331	Sequence 331, App
45	95.2	5.6	1797	9	US-10-174-582-331	Sequence 331, App

ALIGNMENTS

RESULT 1

US-09-745-763-198

; Sequence 198 Application US/09745763

; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

LaVallie, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS: ENCODING THEM

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/09/745,763

APPLICATION NUMBER: 18-Jun-2000

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 198:

SEQUENCE CHARACTERISTICS:

LENGTH: 2930 base pairs

TYPE: nucleic acid


```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match          99.2%; Score 1691.4; DB 10; Length 2930;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1  GTCACCAATCTGCAGCAAGTACCAGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db 598 GTTCCAAATCTGCAGCAAGTACCAGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 67
QY 61  TCTCATAAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAAGCAGAGG 120
Db 658 TCTCATAAATGTGTCATCATGAA-CTCAACAACCTGAACCTGACCCAGGTGCAAGCAGAGG 716
QY 121 AACCTCATCAGAAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC 180
Db 717 AACCTCATCAGAAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC 776
QY 181 AAGAACGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACGCGATTGG 240
Db 777 AAGAACGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACGCGATTGG 836
QY 241 CTGAAGGAGAAAGTGCAGAGCTTTGCAGAGCTGGCTGCCAACAACCTCTGCGTTGGCCAAA 300
Db 837 CTGAAGGAGAAAGTGCAGAGCTTTGCAGAGCTGGCTGCCAACAACCTCTGCGTTGGCCAAA 896
QY 301 GCCAACAAGACACCTTGAGGATGATGACAGCCAGCTCAACTCATTTCACAGTCCAGATG 360
Db 897 GCCAACAAGACACCTTGAGGATGATGACAGCCAGCTCAACTCATTTCACAGTCCAGATG 956
QY 361 GAGAACATCACCACCTATCTCTAAGCCCAAGCAGCAGCAACCTGAAAGACCTTGCAGGACTTA 420
Db 957 GAGAACATCACCACCTATCTCTAAGCCCAAGCAGCAGCAACCTGAAAGACCTTGCAGGACTTA 1016
QY 421 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 480
Db 1017 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 1076
QY 481 CTCCTTTGACAGCGATATTGTGAACATCATTAGCAATATCAGTTTACAGCCCAACACCTG 540
Db 1077 CTCCTTTGACAGCGATATTGTGAACATCATTAGCAATATCAGTTTACAGCCCAACACCTG 1136
QY 541 CGGACCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCCCTTACCAA 600
Db 1137 CGGACCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCCCTTACCAA 1196
QY 601 CACACAGATGATCGACCTCTTGAATAATACCTGGCCCAACATCGGTTTGGATTCTGTT 660
Db 1197 CACACAGATGATCGACCTCTTGAATAATACCTGGCCCAACATCGGTTTGGATTCTGTT 1256
QY 661 TCTCTCAGATGCAACAAGATTGATGAGGTGAGGTTAGACACTGAGTACAGTCCAGTCAAGAA 720
Db 1257 TCTCTCAGATGCAACAAGATTGATGAGGTGAGGTTAGACACTGAGTACAGTCCAGTCAAGAA 1316
QY 721 TCAGTGATTATGAAGAAATGAGCTAGTAGCTTCAAGCAATGCTCAGCTCAAGAAAT 780
Db 1317 TCAGTGATTATGAAGAAATGAGCTAGTAGCTTCAAGCAATGCTCAGCTCAAGAAAT 1376
QY 781 TTTTACAATACTACAAGGTCACCGGGCCCGAGGGTCCCAAGAGGTGACAGAGGATCCCAAG 840
Db 1377 TTTTACAATACTACAAGGTCACCGGGCCCGAGGGTCCCAAGAGGTGACAGAGGATCCCAAG 1436
QY 841 GGACCCCTTGGCCCAACTGGCAACAAGGACAGAAAGAGAGAGGAGGAGGCGCTGGACCA 900
Db 1437 GGACCCCTTGGCCCAACTGGCAACAAGGACAGAAAGAGAGAGGAGGAGGCGCTGGACCA 1496
QY 901 CCTGGCCCTCGCGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 960
Db 1497 CCTGGCCCTCGCGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 1556
```

RESULT 2

```

US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
```

```

QY 1681 CTGTCATCTGCATTA 1695
Db 2277 CTGTCATCTGCATTA 2291
```

```

RESULT 3
US - 09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression P
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,3
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,0

```

```

RESULT 4
US-09-924-340-57
; Sequence 57, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: BejanIn, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
; US-09-924-340-57

Query Match          6.5%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 5.6e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

QY 759 GCATGCTCAGCTCATCAAGAAATTTTACAATACTACAAAGTCCACCGGGCCCCAGGGGTCC 818
Db 109 GCAAGAGCAGGGCCCCCAGGATCTGAAGCCCTCCAGGCCCCAGGCCAGCGGGTCC 168
QY 819 AAGAGGTGACAGAGGATCCAGGACCCCTGGCCCAACTGGCAACAAGGGACAGAAAGG 878
Db 169 CAGAGGAGCGAGGACCCCAAGGTACTCCGGTGAGAAGGGCGACAGGATTTTCAAG 228
QY 879 AGAAGAGGGGAGCCCTGGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTGGACCAAGC 938
Db 229 CCAGCCAGGCTTTCCGGGCCACCGGGTCCCTGGATTCCCGAGGCAAAAGTTGGATCACC 288
QY 939 TGGTCCCCCGG-----AGAGCGTGGGCAAGAGTAAAGGCTCCAGGGCCCC 989
Db 289 TGGCCCCACCTGGCCCTCAAGCAGAGAGGCGACGGAAGGATTCGAGGCCCATCAGGCCT 348
QY 990 CAAAGGCTCCGTGGTTCCTGGGAAGCCCGGCCCTCAGGGCCCCCGGCCCTCGGCTTGGATGG 408
Db 349 GCCTGGCTCCCTGGGCCACCGGGACCTCTCTGGGATTCAGGGCCCCCGCGGCTTGGATGG 408
QY 1050 CCCCCGGGCCACACAGGCAAGAGGGGACTCCCGGCCCTCAGGGGCCCTCTGGCTTCCA 1109
Db 409 TTTGGATGGGAAGGATGGCAAGCTGGCTTGGAGGGGGGACCTCTGGCTCTGGTGGCCCCC 468
QY 1110 GGGACTTCAGGCAACCGTTTGGGAGCGCTGGGTGCTTGGACCTCGGGGACTGCCAGGCTT 1169
Db 469 TGGACTCATGGGACCAACCGGCTTTAAGGGAAACAGGACATCTCTGGCTCCAGGACC 528
QY 1170 GCCTGGGTACCAAGCATGCCAGGCCCCAGGGCCCCCGGCCCTCGGCACTGGCGGCGCAGAGGG 588
Db 529 TAAGGTGACTGTGGCAAAACAGGTCCTCTGGCAGCACTGGCGGCGCCTGGCGCAGAGGG 588
Db 589 TGAACCTGGTGCATGGGACCCAG 613

RESULT 5
US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.054.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29

```

```

; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
; US-09-992-600A-57

Query Match          6.5%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 5.6e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

QY 759 GCATGCTCAGCTCATCAAGAAATTTTACAATACTACAAAGTCCACCGGGCCCCAGGGGTCC 818
Db 109 GCAAGAGCAGGGCCCCCAGGATCTGAAGCCCTCCAGGCCCCAGGCCAGCGGGTCC 168
QY 819 AAGAGGTGACAGAGTCCAGGACCCCTGGCCCAACTGGCAACAAGGGACAGAAAGG 878
Db 169 CAGAGGAGCGAGGACCCCAAGGTAACTCCGGTGAGAAGGGCGACAGGATTTTCAAG 228
QY 879 AGAAGAGGGGAGCCCTGGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTGGACCAAGC 938
Db 229 CCAGCCAGGCTTTCCGGGCCACCGGGTCCCTGGATTCCCGAGGCAAAAGTTGGATCACC 288
QY 939 TGGTCCCCCGG-----AGAGCGTGGGCAAGAGTAAAGGCTCCAGGGCCCC 989
Db 289 TGGCCCCACCTGGCCCTCAAGCAGAGAGGCGACGGAAGGATTCGAGGCCCATCAGGCCT 348
QY 990 CAAAGGCTCCGTGGTTCCTGGGAAGCCCGGCCCTCAGGGCCCCCGGCCCTCGGCTTGGATGG 408
Db 349 GCCTGGCTCCCTGGGCCACCGGGACCTCTCTGGGATTCAGGGCCCCCGCGGCTTGGATGG 408
QY 1050 CCCCCGGGCCACACAGGCAAGAGGGGACTCCCGGCCCTCAGGGGCCCTCTGGCTTCCA 1109
Db 409 TTTGGATGGGAAGGATGGCAAGCTGGCTTGGAGGGGGGACCTCTGGCTCTGGTGGCCCCC 468
QY 1110 GGGACTTCAGGCAACCGTTTGGGAGCGCTGGGTGCTTGGACCTCGGGGACTGCCAGGCTT 1169
Db 469 TGGACTCATGGGACCAACCGGCTTTAAGGGAAACAGGACATCTCTGGCTCCAGGACC 528
QY 1170 GCCTGGGTACCAAGCATGCCAGGCCCCAGGGCCCCCGGCCCTCGGCACTGGCGGCGCAGAGGG 1229
Db 529 TAAGGTGACTGTGGCAAAACAGGTCCTCTGGCAGCACTGGCGGCGCCTGGCGCAGAGGG 588
Db 589 TGAACCTGGTGCATGGGACCCAG 613

RESULT 6
US-09-925-302-64
; Sequence 64, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:

```

```

: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA104
: CURRENT APPLICATION NUMBER: US/09/925,302
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05918
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 896
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 64
: LENGTH: 2691
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (2653)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (2667)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc feature
: LOCATION: (2683)
: OTHER INFORMATION: n equals a,t,g, or c
: US-09-925-302-64

```

```

RESULT 7
US-09-925-301-42
; Sequence 42, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301

```

[illegible]

```
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-880-107-3947

Query Match          6.3%; Score 107; DB 10; Length 5086;
Best Local Similarity 51.3%; Pred. No. 1.3e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 789 ACTACAGGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGTCCAGGGACCC 848
   || || || || || || || || || || || || || || || || || || ||
Db 709 ACAGCCCGTGTCTGGTGTGAAGGGTGAACCTGGTCCCTGTGAAATGGAAC 768
   || || || || || || || || || || || || || || || || || || ||
QY 849 TGCCCCAACTGGCAACAGGGACAGAAAGAGAGAGGGGAGCCTGGACCACTGG 908
   || || || || || || || || || || || || || || || || || || ||
Db 769 AGGTCAACAGAGCCGTGGGCTTCTGGTGAGAGAGACGTGTGGTCCCTGG 828
   || || || || || || || || || || || || || || || || || || ||
QY 909 TCGGGGTGAGAGGCCCAATTTGACCACTGCTGCTCCCGGAGAGCGTGGCGG 968
   || || || || || || || || || || || || || || || || || || ||
Db 829 AGCTGTGCTCCGTGGCAGTATGAAGTGTGGTCCCGTGGTCTGCTGCTCC 888
   || || || || || || || || || || || || || || || || || || ||
QY 969 ATCTAAGGCTCCAGGGCCCGCAAGGCTCCGCTGCTTCCCTGGGAGCCCG 1028
   || || || || || || || || || || || || || || || || || || ||
Db 889 GTCTGTGCTCCCTCCAGGCTTCCAGGTCCTCCAGGTCCTGGCCCAAGG 948
   || || || || || || || || || || || || || || || || || || ||
QY 1029 GGGCCCCAGTGGGGACCCAGGCCCCCGGGCCACAGCAAGAGGGACTCC 1088
   || || || || || || || || || || || || || || || || || || ||
Db 949 TGGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008
   || || || || || || || || || || || || || || || || || || ||
QY 1089 TCAGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1148
   || || || || || || || || || || || || || || || || || || ||
Db 1009 CTCGGCCCCGTTGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068
   || || || || || || || || || || || || || || || || || || ||
QY 1149 ACTCGGGGACTGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1208
   || || || || || || || || || || || || || || || || || || ||
Db 1069 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1128
   || || || || || || || || || || || || || || || || || || ||
QY 1209 CGGCCCTCTGCCCCATCAGGAGCGGTGCTGCTGCTGCTGCTGCTGCT 1268
   || || || || || || || || || || || || || || || || || || ||
Db 1129 TGSCCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1188
   || || || || || || || || || || || || || || || || || || ||
QY 1269 GGC 1271
   ||
Db 1189 AGC 1191

RESULT 9
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match          6.3%; Score 107; DB 9; Length 5145;
Best Local Similarity 51.3%; Pred. No. 1.3e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 789 ACTACAAGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGTCCAGGG 848
   || || || || || || || || || || || || || || || || || || ||
Db 712 ACAGCCCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 771
   || || || || || || || || || || || || || || || || || || ||
QY 849 TGGCCCAACTGGCAACAGGGACAGAAAGAGAGAGGGGAGCCTGGACCACT 908
   || || || || || || || || || || || || || || || || || || ||
Db 772 AGGTCAACAGAGGCCCGTGGCTTCTGCTGAGAGAGGACGCTGTTGGTCC 831
   || || || || || || || || || || || || || || || || || || ||
QY 909 TCGGGTGTGAGAGAGGCCCAATTTGGACCACTGCTGCTCCCGGAGAGCG 968
   || || || || || || || || || || || || || || || || || || ||
Db 832 AGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
   || || || || || || || || || || || || || || || || || || ||
QY 969 ATCTAAGGCTCCAGGGCCCGCAAGGCTTCCGCTGCTTCCCTGGGAGAAC 1028
   || || || || || || || || || || || || || || || || || || ||
Db 892 GTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
   || || || || || || || || || || || || || || || || || || ||
QY 1029 GGGCCCCAGTGGGGACCCAGGCCCCCGGGCCACAGCAAGAGGGACTCC 1088
   || || || || || || || || || || || || || || || || || || ||
Db 952 TGTGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011
   || || || || || || || || || || || || || || || || || || ||
QY 1089 TCAGGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1148
   || || || || || || || || || || || || || || || || || || ||
Db 1012 CTCGGCCCCGTTGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071
   || || || || || || || || || || || || || || || || || || ||
QY 1149 ACTCGGGGACTGCCAGGCTTGCCTGGGGTACAGGATGCGAGGCCCAAG 1208
   || || || || || || || || || || || || || || || || || || ||
Db 1072 TGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1131
   || || || || || || || || || || || || || || || || || || ||
QY 1209 CGGCCCTCTGCCCCATCAGGAGCGGTGCTGCTGCTGCTGCTGCTGCT 1268
   || || || || || || || || || || || || || || || || || || ||
Db 1132 TGGCCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191
   || || || || || || || || || || || || || || || || || || ||
QY 1269 GGC 1271
   ||
Db 1192 AGC 1194

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
US-10-044-090-22

Query Match      6.3%; Score 107; DB 12; Length 5432;
Best Local Similarity 51.3%; Pred. No. 1.3e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps

QY 789 ACTTAAAGGTCCACCGGCCCCAGGGTCCAAAGAGTTCACAGAGATCCCAAGGACCCCC 848
   || || || || || || || || || || || || || || || || || || || || ||
Db 1042 ACAGCCCGGTGCTCTCTGTGTGAAGGGTGAACCTGGTCCCTGGTGAATAATGGAATCC 1101

QY 849 TGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACTGGCCCC 908
   || || || || || || || || || || || || || || || || || || || || ||
Db 1102 AGGTCAACAGAGGCCCGTGGGCTTCTCTGGTGAGAGAGGACGTGTTGGTGCCTTGGCCC 1161

QY 909 TGGGGTGAGAGAGGCCCAATTTGGACACAGCTGGTCCCCCCCCGGAGAGCGTGGCGGCAAGG 968
   || || || || || || || || || || || || || || || || || || || || ||
Db 1162 AGCTGGTGGCCCGTGGCAGTGATGGAAGTGTGGGTCCCGTGGGTCTCTGCTGGTCCCATTTGG 1221

QY 969 ATCTAAGGCTCCAGGGGCCCAAAAGGCTCCCGTGGTTCCTCGGAAGCCCGGCCCTCA 1028
   || || || || || || || || || || || || || || || || || || || || ||
Db 1222 GTCTGTGCCCTTCCAGGCTTCCAGGTGGCCCTGGCCCCAAGGCTGAATTTGAGCTGT 1281

QY 1029 GGGCCCCAGTGGGGAGCCACAGGCCCCCGGGGCCACACAGGCAAGAGGACATCCCGGGCCC 1088
   || || || || || || || || || || || || || || || || || || || || ||
Db 1282 TGGTACGGTGGTCTCTGTGTGGTCCCGCGGTCCCCGCTGGTGAAGTGGTCTTCCAGGCCT 1341

QY 1089 TCAGGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCTTGGGGAGCCTTGGGGTGCCTGG 1148
   || || || || || || || || || || || || || || || || || || || || ||
Db 1342 CTCGGGCCCGGTGGACCTCTCTGGTAAATCTCGAGCAAAAGGCCTTACTGGTGCCAAAGG 1401

QY 1149 ACCTCGGGAGACTCCAGGCTTGCCTGGGGTACCAAGGCATGCCAGGCCCCCAAGGGCCCCC 1208
   || || || || || || || || || || || || || || || || || || || || ||
Db 1402 TGCTGTGCCCTTCCCGGGCTTGTCTGGGGCTCCCGGCCCTCCCTGSAACCCGCGTATTTCC 1461

QY 1209 CGGCGCTCTCGGCCATCAGGAGCGTGGTGGCCCTCGCCCTGCAGAAATGAGCAAAACCC 1268
   || || || || || || || || || || || || || || || || || || || || ||
Db 1462 TGGCCCTGTTGGTGGCCGGTGCTACTGGTGCCAGAGGACTTGTGTGTGAGCCCTGGTGCC 1521

QY 1269 GGC 1271
   ||
Db 1522 AGC 1524

RESULT 12
US-10-001-887-33
; Sequence 33, Application US/10001887
; Patent No. US20020155464A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0269
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 4908
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-887-33

```

Query Match 6.3%; Score 106.8; DB 9; Length 4908;
Best Local Similarity 52.9%; Pred. No. 1.4e-18;
Matches 256; Conservative 0; Mismatches 222; Indels 5; Gaps 1;

QY 795 AGGTCACCGGCGCCAGGGTCCAAAGAGTACAGAGTCCAGGAGCCCTGGGCC 854
DB 1603 AGGGCTGGAGGACCCCAAGAGTGCCTGGCATTCGAGTGACCAAGGGCTTAGTGGCT 1662

QY 855 AACTGCAACAAGAGGACAGAAAGAGAGAGGGAGCTGACACACTGGCCCTGGCGG 914
DB 1663 GCGTGGAAACAGGGTCCCAAGGTGAGAGGGACTTCCTGGGGCCATGGAACCCCTGG 1722

QY 915 TGAGAGAGCCCAATTGGACCACTGGTCCCGGAGAGCTGGCGGCAAGGATCTAA 974
DB 1723 ACCAACTGGGCCCCAAGGTTAGCCGGTTTACGGGTGCGCCCTGGAGGACCAAGGGTGC 1782

QY 975 AGGCTCCAGAGGCCCCAAGGCTCCGCTGGTTCCCTGGGAAGCCGCGCCCTCAGGGCC 1034
DB 1783 AGGAGCCCTGGGGCAAGAAAGTGACTTGGGGCTCCCTGGGCAGCCTGGGCTGAGGGTGC 1842

QY 1035 CAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGACTCCCGCGCCCTCAGG 1094
DB 1843 CTCAGGAATCCAGACTCCAGGTTCCAGCTGCGCCCTATTGGCCCCCAAGGCTGCGGG 1902

QY 1095 CCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAG-----CCTGGGGTCCCTGG 1148
DB 1903 CCTGAAGGGGGAACCAAGGCTGCCAGGGCCCCCTGGAGAGGGAGAGCAAGGGAACCTGG 1962

QY 1149 ACCTCGGGACTTCCAGGCTTCCCTGGGTACCAAGCATGCCAGGCCCAAGGGGCCCC 1208
DB 1963 CACGGTCTGGGCCCCAGGGGCCCCCAAGGGTCCCTGGCTCCCTGGAATCACGGGCCCTCC 2022

QY 1209 CGGCTCCTCGGCCATCAGGAGCGGTGTGCGCCCTGGCCCTGCAGAATGAGCCCAACCC 1268
DB 2023 GGGGCTCCCGGGCCCCGGGACCCCTGGTGGCCCTTCGATGAGACTGGCAT 2082

QY 1269 GCGA 1272
DB 2083 CGCA 2086

RESULT 13
US-09-954-531-961
; Sequence 961, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT FILING DATE: 2002-05-02
; PRIOR FILING DATE: 2002-05-02
; PRIOR FILING DATE: 2000-09-18
; PRIOR FILING DATE: 2000-09-20
; PRIOR FILING DATE: 2000-09-20
; PRIOR FILING DATE: 2000-09-20
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 2000-09-22
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 961
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-961

Query Match 6.3%; Score 106; DB 9; Length 2542;
Best Local Similarity 52.2%; Pred. No. 1.6e-18;

Matches 235; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 782 TTACAATACTACAAGTCCACCGGCGCCAGAGGTTCCAAAGAGCTGACAGAGGATCCCAGG 841
DB 837 TCACAGGCCCCGAAGAGGACAGCGGCCCCCAAGGATCAACGGCAAGGATGGGACCCAG 896

QY 842 GACCCCTCGGCCAACTGCAACAAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCCAC 901
DB 897 GCACGCTTGGCATGAAGGCACTGAGGACAGCGCGGACAGCCCGAAGTCCAGGCCACC 956

QY 902 CTGGGCTTGGGTTGAGAGAGCCCAATTGGACACTGGTCCCGCCGAGAGAGCTGGGG 961
DB 957 AGGGCTTACGCGGTGTCAGCCAGCTTGGGACAAAAGAGGAGCCCTGGAGACCAAGGTG 1016

QY 962 GCAAGAGGATCTAAAGCTCCCAAGGCCCCAAAGGCTCCCGTGGTTCCCTCTGGGAAGCCCG 1021
DB 1017 AGCCGGGCCCCGAGGCTTCTTGATTTCTTGTTCCCTCGGAAAGAGGAGAGCCAG 1076

QY 1022 GGCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCACAGGCAAAAGAGGAGTCC 1081
DB 1077 GGCCTCAGGAGAAATTGGTCCCAAGGCTATCATGGACAGAAAGGTGACCAAGGCGAGA 1136

QY 1082 CCGGCTTCCAGGCCCCCTCTGGCTTCCAGGAGCTTCAGGGCACCGTTGGGGAGCCCTGGGG 1141
DB 1137 GGGGCTCAGTGGGGCAACAGGCCCCCTCAGGAAAGGAGGCGCTTAAGGGGAGCAGGGCC 1196

QY 1142 TGCTTGGGACTCGGGGACTTGCAGGCTTCCCTGGGGTACAGGCTATGCCAGGCCCAAGG 1201
DB 1197 CCCCCGAAATTCAGGGCCCCCAAGGCTTGCCAGGCGTCAAGAGGACAGAGGCTCCCCAG 1256

QY 1202 GCGCCCCCGGCCCCCTCTGGCCCCATCAGGAG 1231
DB 1257 GGAAGACCGGGCCCCCGCGCAAAAGTGGGTG 1286

RESULT 14
US-09-964-824A-255
; Sequence 255, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrihan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT FILING DATE: 2001-09-27
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 2000-09-28
; PRIOR FILING DATE: 2000-09-28
; PRIOR FILING DATE: 2000-09-28
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 255
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-255

Query Match 6.3%; Score 106; DB 10; Length 2542;
Best Local Similarity 52.2%; Pred. No. 1.6e-18;
Matches 235; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 782 TTACAATACTACAAGTCCACCGGCGCCAGGTTCCAAAGAGTGCACAGAGGATCCCAGG 841
DB 837 TCACAGGCCCCGAAGAGGACAGCGGCCCCCAAGGATCAACGGCAAGGATGGGACCCAG 896

QY 842 GACCCCTCGGCCAACTGCAACAAGGAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCCAC 901
DB 897 GCACGCTTGGCATGAAGGCACTGAGGACAGCGCGGACAGCCCGAAGTCCAGGCCACC 956

QY 902 CTGGGCTTGGGTTGAGAGAGGCCCAATTGGACAGCTTGGTCCCGCCGAGAGAGCTGGGG 961


```

Db 957 AGGGCTAGCGGTGTGCGAGGCTGCGGACAAAGAGGAGGCGCTGGAGACCAAGGTG 1016
QY 962 GCAAGAGTCTAAAGGCTCCAGGCGCCCAAGAGGTCCCGTGGTTCCTCCCTGGGAAGCCGG 1021
Db 1017 AGCGGGCCCGAGGCGCTTCTGGATCTCTGGTCCCTCGGAAAGAGGAGAGCCAG 1076
QY 1022 GCGCTCAGGCGCCCGAGTGGGAGCCAGGCGCCCGGCGCCCAAGGCAAGAGGAGCTCC 1081
Db 1077 GCGCTCAGGAGAAATTTGGTCCCGAGGATCATGGGACAGAGGAGTACCAAGGCGAGA 1136
QY 1082 CCGGCGCTCAGGCGCTTCTGCTTCCAGGAGCTTCAGGCGACCGTGGGAGCCTGGGG 1141
Db 1137 GGGGTCAGTGGGCAACAGGCGCTTCAGGAGAGGAGGCGCTTAAGGGGAGAGGCGCC 1196
QY 1142 TGCCTGAGCTCGGAGCTGCCAGGCTTGCCTGGGATACAGGCGATGCCAGGCGCCCAAGG 1201
Db 1197 CCGCGGATTCAGGCGCCCAAGGCTTGCCAGGCGTCAAGAGAGACAGGCGCTCCCGAG 1256
QY 1202 GCGCGCGCGCGCTCCTGCGCCCATCAGGAG 1231
Db 1257 GGAAGACCGGCGCGCGGCAAGTGGGTG 1286

```

RESULT 15

US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:

; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 786

; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-786

Query Match 6.2%; Score 105.4; DB 10; Length 5416;
Best Local Similarity 51.1%; Pred. No. 3.6e-18;
Matches 247; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

```

QY 789 ACTAAGGTCCACCGGCGCCAGGCGTCCAAAGAGGTGCACAGAGGATCCAGGAGACCCCG 848
Db 1039 ACAGCCCGGTGCTCCTGGTGTGAAGGTGAACCTGGTGGCCCTGGTGAAGTGAAGTCC 1098
QY 849 TGGCCCACTGCACAGGAGACAAAGAGGAGAGGAGGCGCTGGACACACCTGGCGCC 908

```

```

Db 1099 AGGTCAACAGAGAGCCCGTGGTCTTCTCGTGAGAGAGACGCTGTTGGTGCCCGCTGGTCC 1158
QY 909 TCGGGGTGAGAGAGAGGCGCCCAATTTGGACCACTGTCTCCCGCGAGAGACGCTGGCGGCAAGG 968
Db 1159 AGCTGGTGGCGCTGGAAGTGTGGAAGTGTGGTCCCGTAGGTCTGCTGCTTAAATGG 1218
QY 969 ATCTAAAGCTCCCAAGGCGCCCAAGAGTCTCCCTGGTGGTTCCTCTGGGAAGCGCGGCTCA 1028
Db 1219 GTCTGTGGCGCTTCCAGGTTTCCAGGTCCCGTGGTCCCAAGGCTGAAATTTGGAGCTGT 1278
QY 1029 GGGCGCGCTAGTGGGACCCAGGCGCCCGCGGCGCCACAGGCAAGAGGAGTCCCGGCGCC 1088
Db 1279 TGGTAACGCTGTCTTACTTACTGAGACCCCGCGTCCCGTGGTGAAGTGGGTCTTCCAGGCT 1338
QY 1089 TCAGGCGCGCTCTGCTTCCAGGAGCTTCAGGCGCACCGTTCGGGAGCGCTGGGTGCTGG 1148
Db 1339 CTCGCGCGCGCTTGGACCTCTCTGGTAATCTTGAGCAACAGCGCTTACTGGTCCCAAGG 1398
QY 1149 ACCTCGGGAGTTCGCGAGGCTTCCCTGGGTTACAGGATGCGAGGCGCCCAAGGCGCGCC 1208
Db 1399 TGCTGTGGCGCTTCCCGCGCTTCTGGGCTCCCGGCTCCCTGGACCGCGGCTATTC 1458
QY 1209 CGGCGCTCTGCGCCCATCAGGAGCGGTGGTGGCGCTGGCGCTGCAGAAATGAGCCACCC 1268
Db 1459 TGGCGCTCTGCTGTGCTGCGGTACTACTGTGCGCAGAGGACTTGTGTGAGCGCTGGTCC 1518
QY 1269 GGC 1271
Db 1519 AGC 1521

```

Search completed: March 21, 2003, 08:22:02
Job time : 277.28 secs

Gencore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 2223.7 Seconds
(without alignments)

12344.923 Million cell updates/sec

Title: US-09-763-712a-1_COPY_1_1695

Perfect score: 1695
Sequence: 1 gtccacgaatcgcagcaaga.....cagctactgtcatcgtacatta 1695

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1357.4	80.1	3305	11	BC009162 Mus muscu
2	677.8	40.0	957	14	B0955927 AGENCOURT
3	656.4	38.8	936	14	B0891432 AGENCOURT
4	616.8	36.4	884	14	B0934501 AGENCOURT
5	559	33.0	906	13	B1456109 AGENCOURT
6	556.4	32.8	861	14	B0713873 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
7	525.4	31.0	668	14	B0674807 AGENCOURT
8	488.6	28.8	511	12	BF805750 OVI-C1017
9	480.2	28.3	581	14	B0127513 i160h06.y
10	473.4	28.0	967	9	AL568743
11	445.6	26.3	580	12	BE910803
12	437.2	25.8	552	10	AN958053
13	415.8	24.5	601	10	BE290299
14	395.8	23.4	410	12	BF856923
15	369.4	21.8	638	10	BB248064
16	351.4	20.7	500	14	BM676508
17	350	20.6	507	14	BM713891
18	343.8	20.3	808	14	BQ771366
19	311	18.3	339	9	AA304251
20	301.4	17.8	357	10	AM379436
21	297.2	17.5	517	9	AL543000
22	296	17.5	310	9	AA361740
23	295.2	17.4	679	14	W27610
24	271	16.0	350	9	AA012704
25	258.2	15.2	286	12	BF854324
26	245.8	14.5	420	12	BE929813
27	244.2	14.4	252	13	BI051416
28	244	14.4	277	12	BF854320
29	242.2	14.3	382	12	BE012097
30	241	14.2	550	12	BF858081
31	233.8	13.8	429	12	BE929798
32	230.8	13.6	465	10	AM240221
33	227.6	13.4	930	9	AL541116
34	225	13.3	922	9	AL542672
35	223.2	13.2	280	13	BI010883
36	212.6	12.5	281	12	BF854408
37	212	12.5	723	14	BO004176
38	210.2	12.4	943	14	BO087163
39	209.8	12.4	628	14	BM721855
40	205.8	12.1	261	13	BI010880
41	205.6	12.1	1082	14	BM907108
42	201	11.9	655	12	BG482931
43	200.8	11.8	345	9	A1742661
44	199.6	11.8	441	10	BB749743
45	193.4	11.4	234	12	BF854188

ALIGNMENTS

RESULT 1
LOCUS BC009162 3305 bp mRNA linear HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to scavenger receptor with C-type lectin,
clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3305)
REFERENCE Strausberg, R.
TITLE Direct Submission
AUTHORS Submitted (05-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
CONTACT: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu
Gunnarone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLINL at: <http://image.llnl.gov>
Series: IRK Plate: 5 Row: 5 Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

Location/Qualifiers

Source

1..3305

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="C57BL/6J"

/clone="IMAGE:2811487"

/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months

old, gross tissue."

/clone_lib="NCI_CGAP_Mam5"

/lab_host="DH10B"

/note="Vector: PCMV-SpORF6"

BASE COUNT 1031 a 757 c 760 g 757 t

ORIGIN

Query Match 80.1%; Score 1357.4; DB 11; Length 3305;

Best Local Similarity 87.6%; Pred. No. 0;

Matches 1484; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

```
QY 1 GTCACGATCTGCAGACAGATACCGCGTCTCCAGGCAATCTGCAGAACCAATGTAT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 627 GTCACCAATCTGCACAAAGATAGTGTCTCCAGGCAATCTGCAGACCAATGTAT 686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TCTCATATGTGTCATCATCAAGCAACCTGAACCTGACCCAGGCGTACAGAGG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 687 TCTAGAGTGTGTTATCATGAACTCAACCTGAACTGACCCAGGTTGACAGAGG 746
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 AACCTCATCAAGATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCATC 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 747 AACCTTATCTCAATTTGTCAGACAGTGTGTGGATGACACAAAGCCAGCTACAGCATT 806
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 AAGACGACTTTTCAAAATCTGCAGAGGTTTTTTTCAAGCCAAAGAGACAGCATTTGG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 807 AAGAAATATTTCACAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGACAGCATTTGG 866
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 CTGAAGAGAGAAAGTGCAGAGCTTGCAGACGCTGCTCCAAACATCTGTGGTTGGCCAAA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 867 CTAAAGAGAAAGATGACAAAGCTTGCAGACATTTGGCTGCCAAACATCTGTGGCCAAA 926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GCCAACACAGCACCCCTGAGAGATATGACAGCCAGCTCAACTCATTCACAGAGTCAGATG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 927 GCCAACATGACACCTTAGAGATATGATATGACAGCTCAGCTCATTCACAGAGTCAGATG 986
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 GAGACATCTCACTATCTCTCAAGCCAGAGAGAGAACTGAAAGCTTGCAGAGATTTA 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 987 GACAACTTTCACCTATCTCAAGCCAGAGAGAGAGAGCTTGAAGAGCTTGCAGAGATTTA 1046
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 CACAAAGATGAGAGATGAGACAGCCATCAAGTTCAACCAAGTGAAGAGAGGTTTCCAG 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1047 CACAAAGATGAGAGAAATGAGACAGCTGTCAAGTTCAAGCTTGAAGAGAGGTTTCCAG 1106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 CTCTTTGAGAGGATATTTGAAACATCATTTAGCAATATCATGTTACACAGCCCAACCATG 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1107 GTCTTTGAGACAGATATTTGAAACATCATTTAGCAACATTCAGCTCAGCCCATCATCTG 1166
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 CGGAGCTGACAGCAATCTTAATGAAGTACAGACACTTGCACAGATACCTTTACCAAA 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1167 CGGACACTGACCAAGCAATCTGAATGATGAGACCAACATGACAGACACTTGCACAGA 1226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 CACACAGATGATGACCTCTGGAATATACCTGCGCAACATCCCTTTGGATTTCTGT 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1227 CACGAGATGACCTCTCTGGAATATACCACTAGTCAACATCCCTTTGGATTTCTATT 1286
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 661 TCTCTCAGAGTGCACAGATTTTGATAGGTGAGGTTAGACACTGGAATAGCCCACTTA 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1287 TCTCTCAGAGTGCACAGACATGATAGGTGGAAGTTAGACACTGGAATAGCCCACTTA 1346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 721 TCACTGATTTAGAGAAATGAAAGCTAGTACCTCCAGAGATGTCAGTCTATCAAGAT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1347 TCACTGATTTAGAGAAATGAAAGCTAGTACCTCCAGAGAGCTGCTATCAAGAT 1406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 TTTCAATTAATACAGAGTCCACCGGCTCCAGAGGCTCCAGAGATGACAGAGATCCAG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1407 TTTTACATTTTACAGAGTCTCTCTGAGGCTCCAGAGATGACAGAGATCCAG 1466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 GGACCCCTGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1467 GGACCCCTGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 CTTGAGGCTTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1527 CTTGAGGCTTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 1586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 GGCAAGAGATCTAAAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCC 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1587 AGCAAGAGATCTAAAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 1646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 GGCCCTCAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1647 GGCCCTCAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1706
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 CCGGAGGCTTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1707 CCGGAGGCTTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 1766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1141 GTGCTGAGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1767 GTGCTGAGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCT 1826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1201 GGCCCTCAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1827 GGACCACTGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1261 CCAACCCGCGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1887 CCAACCCGCGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAGGCTCCAGAG 1946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1321 TGCCTATATTTTCACTGAGAAAGATTTTGAGAGATGCAAGCTTTCTGTGAAGAC 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1947 TGCCTATATTTTCACTGAGAAAGATTTTGAGAGATGCAAGCTTTCTGTGAAGAC 2006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1381 AAGCTTTCACATCTTGTTCATTAACCTAGAGAGAGCAATGATTAAGAAACAG 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2007 AAGCTTTCACATCTTGTTCATTAACCTAGAGAGAGCAATGATTAAGAAACAG 2066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1441 ATGCTAGAGAGAGAGAGCACTGAGTGGCTCCAGAGCTCAAGAGCTGAAGATG 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2067 ACCGTGGGAGAGAGAGCACTGAGTGGCTCCAGAGCTCAAGAGCTGAAGATG 2126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1501 AAGTGGCTGAGTGGAGATCTCCAGAGCTCAAGAAATTTGAAGAGCTGAGAGCAAT 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2127 AAGTGGCTGAGTGGAGATCTCCAGAGCTCAAGAAATTTGAAGAGCTGAGAGCAAT 2186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1561 TGGGCTATAGGCTATGAGGCTGAGAGAGAGTGTGCTGGTGTATTTATGCTGGGAGTGG 1620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2187 TGGGCTATAGGCTATGAGGCTGAGAGAGAGTGTGCTGGTGTATTTATGCTGGGAGTGG 2246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1621 AAGGATTTCCAAATGTAAGAGCTCAATTAATTTTGCAGAGAGAGAGAGAGTGA 1680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2247 AAGGATTTCCAAATGTAAGAGCTCAATTAATTTTGCAGAGAGAGAGAGAGTGA 2306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1681 CTGTCATCTGATTA 1695
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2307 CCATTCATCATTAATTA 2321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 2
BO955927          957 bp      mRNA      linear      EST 21-AUG-2002
LOCUS             BO955927
DEFINITION       AGENCOURT.8763247 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313500
5', mRNA sequence.
ACCESSION        BO955927
VERSION          BO955927.1 GI:22371405
KEYWORDS         EST.
SOURCE           house mouse.
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE        1 (bases 1 to 957)
AUTHORS          NIH-MGC http://mgc.nci.nih.gov/.
TITLE            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL          Unpublished (1999)
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cga@bbs-remail.nih.gov
                  Tissue Procurement: Susan L. Sullivan, Ph.D.
                  cDNA Library Preparation: Resgen, Invitrogen Corp
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLAM13740 row: 1 column: 13
                  High quality sequence stop: 665.
FEATURES
    source
        1..957
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="IMAGE:6313500"
        /lab_host="NIH_MGC_129"
        /note="Organ: Olfactory epithelium; Vector:
        PCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned
        unidirectionally. Primer: Oligo dT. Average insert size
        2.2 kb. Constructed by: Resgen, Invitrogen Corp. Note: this
        is a NIH_MGC library."
BASE COUNT       260 a 269 c 258 g 169 t 1 others
ORIGIN
Query Match      40.0%; Score 677.8; DB 14; Length 957;
Best Local Similarity 86.9%; Pred. No. 2.3e-155;
Matches 780; Conservative 0; Mismatches 113; Indels 5; Gaps 3;
OY 330 CAGCAGCTCAATCATTCACAGTCAGATGAGAACATCACCACTATCTCAAGCCAA 389
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 13 CCGGATCTCAGCTCATTCACAGTCAGATGAGAACATCACCACTATCTCAAGCCAA 72
OY 390 CAGCAGAACCTGAAAGACCTGCAGACTTACCAAAAGTGCAGAGATAGAAGACCCAT 449
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 73 CAGAGACCCCTGAAAGACCTTACAGACTTACCAAAAGTGCAGAGATAGAAGACCCAT 132
OY 450 CAGTTCACCACTGAGAGAGAGCTTCCAGCTTTGAGAGGATATGTGAACATCAT 509
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 133 CAGTTCACCACTGAGAGAGAGCTTCCAGCTTTGAGAGATATGTGAACATCAT 192
OY 510 TAGCAATATCACTTACACAGCCACCCTGCGAGAGCTGACCAAGCATCTAAATGAAGT 569
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 193 TAGCAATATCACTTACACAGCCACCCTGCGAGAGCTGACCAAGCATCTGATGATGT 252
OY 570 CAGGACCACTTGCAGAGATACCTTACCAAAAGTGCAGAGATAGAAGACCCAT 629
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 253 TAGGACCACTTGCAGAGATACCTTACCAAAAGTGCAGAGATAGAAGACCCAT 312
OY 630 TACCTTGAGCAATCCGTTTGATTTCTCTCTCAGAGATGAAGATTTGATGAG 689
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 313 CACACTAGTCAACATCCGCTTGATTTCTCTCTCAGAGATGAAGATGAAGTGAAG 372
OY 690 GTGAGGTTAGCACTGAGATGAGCAACTTATCAGTATGAGAAATGAAGCTAGT 749
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

```

DB 373 GTCAAGTTAGCACTGAGATGAGGCCAATTCAGTGTATGAGAGATGAAGATCGT 432
OY 750 AACTTCAGCATGTGTACCTCATCAAGATTTTACATACATGAGTCCACGGGCC 809
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 433 TCACTTCAGCAGCGGTACCTCATCAAGATTTTACATACATGAGTCCACGGGCC 492
OY 810 CAGGGTCCAAAGAGTGCAGAGATCCAGGAGACCCCTGCGCCCACTGAGCAAGG 869
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 493 CAGAGTCCAAAGAGTGCAGAGATCTCAGGAGACCCCTGCTCACTGAGCAAGG 552
OY 870 ACAGAAAGAGAGAAAGGGAGCTTGAGACCTGAGCCCTGCGGTGAGAGAGCCCAAT 929
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 553 ACAGAAAGAGAGAAAGGGAGAGAGCTTGAGACCTGAGCCCTGCGGTGAGAGAGCCCAAT 612
OY 930 TGCAGCACTGTGTCCCGGAGAGAGCTGCGGCAAGAGATCTAAAGCTTCCAGGGCC 989
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 613 TGCAGCACTGTGTCCCGGAGAGAGCTGCGGCAAGAGATCTAAAGCTTCCAGGGCTC 672
OY 990 CAAAGCTCCGTTGTTCCCTGAGAGCCCGGCTCAGAGGCCCTGAGGGAGACCCAGG 1049
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 673 CAAAGATCTCTGTGGTCCCGCAGAGAGAGCTTGAGCCCTCAAGAGCTAGTGGAGACCCAGG 732
OY 1050 CCCCCCGGGCCAGCAGGAAAGAGGAGTCCCGGCTCAGAGGCCCTCTGCTTCA 1109
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 733 ACCACAGGTCCAGCAGGCAAGAGATGAGTCCCTGAGCCCTCANGGCCCTCTGAGCTT -CA 791
OY 1110 GGGACTTCAGGGCAGCGTGTGGGAGCTGGGCTGAGTCTCGGGAGCTG -CCAGGCT 1168
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 792 GGGACTTCAGGGCAGCGTGTGGGAGCTGGGCTGAGTCTCGGGAGCTG -CCAGGCT 851
OY 1169 TGGCTGGTACAGCATATGCC---AGGCCCAAGGGCCCGGCGCTCTGAGGCC 1223
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 852 TGGCAGGGTGGCAGCATATGCCCTGGGGCTTAAGGAGCACCACCTGGGCCCTCCAGAGCC 909
RESULT 3
BO891432          936 bp      mRNA      linear      EST 16-AUG-2002
LOCUS             BO891432
DEFINITION       AGENCOURT.8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
5', mRNA sequence.
ACCESSION        BO891432
VERSION          BO891432.1 GI:22283446
KEYWORDS         EST.
SOURCE           house mouse.
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE        1 (bases 1 to 936)
AUTHORS          NIH-MGC http://mgc.nci.nih.gov/.
TITLE            National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL          Unpublished (1999)
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Email: cga@bbs-remail.nih.gov
                  Tissue Procurement: Mark Macdonochie, Ph.D. and Nancy L. Freeman,
                  Ph.D.
                  cDNA Library Preparation: Resgen, Invitrogen Corp
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Agencourt Bioscience Corporation
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLAM13790 row: n column: 19
                  High quality sequence start: 25
                  High quality sequence stop: 632.
FEATURES
    source
        1..936
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="IMAGE:6332754"
        /lab_host="NIH_MGC_130"
        /note="Organ: Olfactory epithelium; Vector: PCMV-SPORT6.1.ccd;
        Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally."

```

Primer: Oligo dt. Average insert size 1.95 kb.
 Constructed by Resgen, Invitrogen Corp. Note: this is a
 NIH_MGC Library."

BASE COUNT 273 a 250 c 244 g 163 t 6 others
 ORIGIN

Query Match 38.8%; Score 658.4; DB 14; Length 936;
 Best Local Similarity 85.9%; Pred. No. 1.3e-150;
 Matches 752; Conservative 0; Mismatches 119; Indels 4; Gaps 2;

QY 227 AGGACACGGGTTGGCTTGAAGAGAAAGTGAAGAGCTTGACAGCCCTGGCCGACACAACT 286
 DB 33 AAGAACCGATGGCTTAAGAGAAAAAATGACAGACTTGACAGACTGGCTGCCAAGCACT 92
 QY 287 CTCGGTGGCCAAAGCCAAAGACACACCTGGAGATGACAGACGACCTCACTACAT 346
 DB 93 CTGCCCCGGCCAAAGCCAAAGATGACACCTAGAGATGATGAATAGCCACTGCTCAT 152
 QY 347 TCACAGGTGAGATGAGACATCACACTATCTCTCAAGCCAAAGACAGACAGACCTGAAG 406
 DB 153 TCACAGGTGAGATGAGACATCACACTATCTCTCAAGCCAAAGACAGACAGACCTGAAG 212
 QY 407 ACCTGAGGACTTACCAAGATGACAGATGACAGACGACCTCACTACATGACCACTG 466
 DB 213 ACCTGAGGACTTACCAAGATGACAGATGACAGACGACCTCACTACATGACCACTG 272
 QY 467 AGGACGCTTCCAGCTTTGAGAGAGATTTGAGACATCATTAAGCAATATGACATGAC 526
 DB 273 AGGACGCTTCCAGCTTTGAGAGAGATTTGAGACATCATTAAGCAATATGACATGAC 332
 QY 527 CAGCCCAACCTGCGGAGCCTGACAGCAATCTAAATGAATGACAGCACTTGACAG 586
 DB 333 CAGCCCAACCTGCGGAGCCTGACAGCAATCTAAATGAATGACAGCACTTGACAG 392
 QY 587 ATACCCCTTACCAACACAGATGATGACCTGCTGATGATGATGATGATGATGATGATG 646
 DB 393 ATACCCCTTACCAACACAGATGATGACCTGCTGATGATGATGATGATGATGATGATG 452
 QY 647 GTTGGATTTCTTCTCTCAGATGACAGATGATGATGATGATGATGATGATGATGATG 706
 DB 453 GTTGGATTTCTTCTCTCAGATGACAGATGATGATGATGATGATGATGATGATGATG 512
 QY 707 AAGTACCACTTATCTAGTATGAGAAATGAAGCTATGATGATGATGATGATGATGATG 766
 DB 513 AAGTACCACTTATCTAGTATGAGAAATGAAGCTATGATGATGATGATGATGATGATG 572
 QY 767 AGCTATCAAGATTTTACATATCAAGTCCACCGGCCCCAGGGGTCCAAAGAGGTG 826
 DB 573 AGCTATCAAGATTTTACATATCAAGTCCACCGGCCCCAGGGGTCCAAAGAGGTG 632
 QY 827 ACAGAGATCCCAAGGACCCCTGCCCCAATGCGCAACAAGGACAGAAAGAGAGAG 886
 DB 633 ACAGAGATCCCAAGGACCCCTGCCCCAATGCGCAACAAGGACAGAAAGAGAGAG 692
 QY 887 GGAAGCTGACACCTGCGGCTGCGGCTGAGAGAGGCCCAATGACACAGCTGCTCCC 946
 DB 693 GGAAGCTGACACCTGCGGCTGCGGCTGAGAGAGGCCCAATGACACAGCTGCTCCC 752
 QY 947 CCGAGAGCTGCGGCAAGAGATCTAAAGCTCCAGGCCCCCAAGG--CTCCCTG 1004
 DB 753 CCGAGAGCTGCGGCAAGAGATCTAAAGCTCCAGGCCCCCAAGGCTCTCGTGGG 812
 QY 1005 TTCCCTGGGAGAGCCCGCCCTCAAGGCCCCAGTGGGGACCCAG--CCGCCGGGCCCA 1062
 DB 813 TTCCCTGGGAGAGCCCGCCCTCAAGGCCCCAGTGGGGACCCAGGNNACCAAGGCTCCAC 872
 QY 1063 CAGGCAAGAGAGAGCTCCCGCCCTCAAGGCCCC 1097
 DB 873 CAGGCAAGAGAGAGCTCCCGCCCTCAAGGCCCC 907

RESULT 4
 BQ934501

LOCUS BQ934501 884 bp mRNA linear EST 21-AUG-2002
 DEFINITION AGENCOURT_8765685 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330794
 5', mRNA sequence.
 ACCESSION BQ934501
 VERSION BQ934501.1 GI:22349884
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 884)
 NIH-MGC <http://mgi.mc.nhl.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgrabs-remail.nih.gov
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
 Ph.D.
 cDNA Library Preparation: Resgen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLNL3785 row: m column: 03
 High quality sequence stop: 352.
 Location/Qualifiers
 1..884
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:6330794"
 /clone_lib="NIH_MGC_130"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: oocytes; Vector: pCMV-Sport6.1.cdb;
 Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
 Primer: Oligo dt. Average insert size 1.95 kb.
 Constructed by Resgen, Invitrogen Corp. Note: this is a
 NIH_MGC Library."

BASE COUNT 282 a 225 c 205 g 172 t
 ORIGIN

Query Match 36.4%; Score 616.8; DB 14; Length 884;
 Best Local Similarity 84.2%; Pred. No. 2.1e-140;
 Matches 743; Conservative 0; Mismatches 132; Indels 7; Gaps 4;

QY 49 AACCAATGTATCTCATATGTCATCATGATCACTCAACCACTGACCTGACCCAG 108
 DB 1 AGCCAAATGTATCTCATATGTCATCATGATCACTCAACCACTGACCCAG 60
 QY 109 GTGCAGCAGAGAACTCATCAAGATCTGACGCTGTGTGATGACACAGAGCT 168
 DB 61 GTTCAGCAGAGAACTCATCAAGATCTGACGCTGTGTGATGACACAGAGCT 120
 QY 169 ATCCAGGATTCAGAAAGCCTTCAAAATCTCAGAGGTTTCTTCAAGCCAGAG 228
 DB 121 ATCCAGGATTCAGAAAGCCTTCAAAATCTCAGAGGTTTCTTCAAGCCAGAG 180
 QY 229 GACACGATTTGGCTGAAGAGAGAGTGCAGAGCTTGCAGAGCTGCTGCCAACAATCT 288
 DB 181 GACACGATTTGGCTGAAGAGAGAGTGCAGAGCTTGCAGAGCTGCTGCCAACAATCT 240
 QY 289 GCGTTGGCCAAAGCCAAACAGACACCTGAGAGATGTAAGACAGCCAGCTCAATTC 348
 DB 241 GCGTTGGCCAAAGCCAAACAGATGACACCTGAGAGATGTAAGATGACAGCTCAATTC 300
 QY 349 ACA-GGTCAATGAGAGACATCACTATCTCAAGCCAAAGAGAGAGAGAGAG 407
 DB 301 ACAAGGTCATATGAGAGACATCACTATCTCAAGAGCCAAAGAGAGAGAGAG 360
 QY 408 CCTGCAGACTTACACAAAGATGACAGATGAGAGAGAGAGAGAGAGAGAGAGAG 467
 DB 361 CTTTCAGAGCTTACACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

QY 468 GGACGCTTCAGCTCTTTGAGAGCATATTTGTGAACATCATTTAGCAATATTCAGTTACAC 527
 |||||
 Db 421 GGAAGCTTCAGGCTTTGAGACAGATTTGTGAACATCATTTAGCAATATTCAGTTACAC 480
 |||||
 QY 528 AGCCCAACCTGCGGAGCTGACACGACATCTAATGAAGTCAGAGCATTGACAGCA 587
 |||||
 Db 481 AGCCCAACCTGCGGAGCTGACACGACATCTAATGAAGTCAGAGCATTGACAGCA 540
 |||||
 QY 588 TACCTTACCAACACAGATGATGTCCTGCTTGAATATTTACCTGCGCAACATCCG 647
 |||||
 Db 541 CACCTTGACACACACAGGATGACCTGACCTCTTGAAT -ACACACTGATCAACATCCG 599
 |||||
 QY 648 TTTGATTTCTTTCTCTCAGAGTCAACAGATTTGATGAGTGTGAGTGTGAGACTGA 707
 |||||
 Db 600 CTGATTTCTATTTCTCTCAGAGTCAACAGATTTGATGAGTGTGAGTGTGAGACTGA 659
 |||||
 QY 708 AGTACCAACTATTCAGTATTTATGAGAGAAATGAAGCTAGTACTCCAAAGCTGTA 767
 |||||
 Db 660 AGTGCACCACTTATCAGTGTATGAGAGATGAAGATGTAAGCTGTAAGCTCCAGCAGTCA 719
 |||||
 QY 768 GCTATCAAGATTTTACAACTACTCAAGTCCACCGGCGCCAGGAGGTCGAAGAGTG 826
 |||||
 Db 720 GCTCATCAAGACTTATACCTCTTACAGGTCCTCTGCGCCCAAGAGTCAAAAGAGG 779
 |||||
 QY 827 --ACAGAGGATCCAGGAGACCCCTGG--CCCACTGGCAGACAGAGGACAGAAAGAG 882
 |||||
 Db 780 ACAGAGGATCTCAGGAGGACCAACTGGGTCCAACTTGGCAGACAGAGGAGAAAGAGG 839
 |||||
 QY 883 AAGGGGAGGCTGACGACCTGCGCTGGGGTGAAGAGGC 924
 |||||
 Db 840 AAGAGAAAGGAGAGGAGGCTGTGTGCTCCACCTGCGGCGCTGCGC 881
 |||||

RESULT 5
 B1456109 906 bp mRNA linear EST 21-AUG-2001
 LOCUS 603172765P1 NCI_CGAP_Mam5 Mus musculus cDNA IMAGE:5251888 5',
 DEFINITION mRNA sequence.
 B1456109
 ACCESSION B1456109.1 GI:15246765
 VERSION EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 906)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@bhs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM11636 row: b column: 17
 High quality sequence stop: 730.
 Location/Qualifiers
 1. 906
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="5251888"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators

FEATURES

source
 1. 906
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="5251888"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators

BASE COUNT 247 a 228 c 256 g 175 t
 ORIGIN
 Query Match 33.0%; Score 559; DB 13; Length 906;
 Best Local Similarity 82.4%; Pred. No. 3.2e-126;
 Matches 703; Conservative 0; Mismatches 140; Indels 10; Gaps 5;

QY 548 TGACCAAGCATCTAATATGAGTCAGAGCATTGCGACATATCCCTTACCAGACAGAG 607
 |||||
 Db 1 TGACCAAGCATCTAATATGAGTCAGAGCATTGCGACATATCCCTTACCAGACAGAG 60
 |||||
 QY 608 ATGATCTGACCTCCTTGAATATTTACCTGCGCAACATCCGTTGATCTGTTCTCTCA 667
 |||||
 Db 61 ATGACCTGACCTCCTTGAATATTTACCTGCGCAACATCCGTTGATCTTATTTCTCTCA 120
 |||||
 QY 668 GATGCAACAGATTTGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 727
 |||||
 Db 121 GATGCAACAGATTTGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 180
 |||||
 QY 728 TTATGGAAGAAATGAAGCTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 787
 |||||
 Db 181 TTATGGAAGAAATGAAGCTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 240
 |||||
 QY 788 TACTACAAAGTCCACCGGCGCCAGGAGGTCGAAGAGTGTGAGTGTGAGTGTGAGTGTGAG 847
 |||||
 Db 241 TTCTACAAAGTCTCTCTGCGCCCAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 300
 |||||
 QY 848 CTGGCCCAACTGCGCAACAGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 907
 |||||
 Db 301 CTGGCCCAACTGCGCAACAGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 |||||
 QY 908 CTGGCCCAACTGCGCAACAGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
 |||||
 Db 361 CTGGCCCAACTGCGCAACAGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 |||||
 QY 968 GATCTAAAGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1027
 |||||
 Db 421 GATCTAAAGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
 |||||
 QY 1028 AGGGCCCAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1087
 |||||
 Db 481 AAGGAGCTTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
 |||||
 QY 1088 CTGAGGCGCTCTCTGCTTCCAGGAGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1147
 |||||
 Db 541 CTGAGGCGCTCTCTGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
 |||||
 QY 1148 GACCTGGGAGCTGCGCAGGCTTGGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1207
 |||||
 Db 601 GACCTGGGAGCTTGGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660
 |||||
 QY 1208 CCGGCGCTCTCTGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1267
 |||||
 Db 661 TTAGG -CCTCAGAGGCGCTCAGAGGAG -ATGAGAGCAATTTGTGACAGAAATGAACAAGCC 718
 |||||
 QY 1268 CCGGCGGAGGAGCAATGGC--TGCCGCGCTCACTGTGAAGAACTTACA--GACAATG 1322
 |||||
 Db 719 CAGCAATCAGAGGCTTCAACCGATGTCCTCAGTGAAGAACTTACA--GACAATG 778
 |||||
 QY 1323 CTACTATTTTTCAGTGTG--AGAAAGAAATTTTGTGAAGTGAACCTTTCTGTGAGA 1379
 |||||
 Db 779 CTACTATTTTTCATTTGACCAAAAGAAATTTTGTGAAGTGTGCTAAGCGTTCTGTGAGA 838
 |||||
 QY 1380 CAAGTCTTCACAT 1392
 |||||
 Db 839 AAAATTTTTCACAT 851
 |||||

RESULT 6
 B0713873 861 bp mRNA linear EST 16-JUL-2002
 LOCUS B0713873

	DEFINITION	AGNCOURT_8492828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240 sequence.					
	ACCESSION	U713873					
	VERSION	B0713873.1 GI:21852772					
	KEYWORDS	EST.					
	SOURCE	house mouse.					
	ORGANISM	Mus musculus					
	REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalacostraca; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
	AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/.					
	TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
	JOURNAL	Unpublished (1999)					
	COMMENT	Contact: Robert Strausberg, Ph.D. Email: csapbs@mail.nih.gov Tissue Procurement: Susan L. Sullivan, PhD. cDNA Library Preparation: Resgen, Invitrogen Corp CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLM13721 row: n column: 01 High quality sequence stop: 598.					
	FEATURES	Location/Qualifiers					
	Source	1..861 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:6306240" /clone_id="NIH_MGC_129" /lab_host="DH10B (phage-resistant)" /note="Organ: olfactory epithelium; Vector: pCMV-Sport6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: This is a NIH MGC Library."					
	BASE COUNT	221 a 253 c 239 g 147 t 1 others					
	ORIGIN						
	Query Match	32.8%; Score 556.4; DB 14; Length 861;					
	Best Local Similarity	82.8%; Pred. No. 1.4e-125;					
	Matches 647; Conservative	0; Mismatches 132; Indels 2; Gaps 1;					
OY	683	TGATGAGTCGCGAGTTAGCACTGAAGTATGCCAATTATCACTGATTTATGAGAATAA	742				
DB	14	TGATGAGTCGCGAGTTAGCACTGAAGTATGCCAATTATCACTGATTTATGAGAATAA	73				
OY	743	AGCTAGTAGACTCCAAAGCAGTGTCAGCTCATCAAGAAATTTTACAATTA	802				
DB	74	AACTGTGTGACTCCAAAGCAGTGTCAGCTCATCAAGAAATTTTACAATTA	133				
OY	803	GCGGGCCCCAGGGGTCGAAGAGTGCACAGAGATCCAGGAGACCCTGGCCCACTGGCA	862				
DB	134	CTGGGCCCCAGAGGTCGAAAAGGTGACAGAGATCTCAGGAGCACCTGGTCCA	193				
OY	863	ACAAGGAGCAGAAAAGAGAGAGAGGAGGAGCCCTGAGACCACTGGCCCTGGGTGAGAG	922				
DB	194	ACAAGGAGCAGAAAAGAGAGAGAGGAGGAGCCCTGAGACCACTGGCCCTGGGTGAGAG	253				
OY	923	GCCCAATTGAGCACTGGTCCCCTGGGAGAGCGTGGGCGCAAAGATCTTAAGCTCC	982				
DB	254	GCACAAATTGAGCACTGGTCCCCTGGGAGAGCGTGGGCGCAAAGATCTTAAGCTCC	313				
OY	983	AGGGCCCCCAAGAGGTCGCTGGTGGTCCCTGGGAGAACCCGGGCCCTGAGGAGCCCA	1042				
DB	314	AGGGTCCTCAAGAGGATCTGCTGGGTCCTCCAGGAGAACCTGGCCCTCAAGGACCTA	1102				
OY	1043	ACCCAGAGGCCCCCGGGCCACCAGGCAAGAAGAGACTCCCGGACCTCAGGAGCCCTCTG	1102				
DB	374	ACCCAGAGGACACCAAGTCCACCAAGGCAAGAAGATGACCTCCCTGAGGAGCCCTCTG	433				
OY	1103	GCTTCAGGAGACTTACGGGACACCGTTGGGAGACCTGGGGTGCCTTGACCTGGGAGCTGC	1162				

DB	434	GCTTCAGGAGCTACAGGGCACTGTGGGGAGCCTGTGAGTACTGTGAGACTCGGGGTTGC	493
QY	1163	CAGGCTGCTCTGGGTAGCAGGCATGCCAGGCCCCCGGCGCTCTGTGGCC	1222
DB	494	CAGGCTTGCCAGGAGGTGGCCAGGCAATGCCCTGGGGCTTAAGGAGACCCTGGCTCCAGGCC	553
QY	1223	CATCGAGAGAGGGTGGTGGCCCTGGGCGCTGCATAATGAGGCAACCCCGGAGCGGAGACA	1282
DB	554	CCTCAGGAGCAATGAGAGCATTTGGCTGTCTGCATAATGAAACCAACCCCAACATCAGAGTCA	613
QY	1283	ATGGCTGCCCGCTCAGTGGAAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGA	1342
DB	614	ACGGATGTCGCCGCTCAGTGGAAAGAACTTCACAGATAAAGCTACTATTTTTCCTGGCAA	673
QY	1343	AAGAATTTTTCAGAGTGCAGAAC--TTTTCGTGAGAGCAAGCTTTCACATCTGTGTT	1400
DB	674	AAAAAATTTTCGAAATATCTACCTTTTTCGTGCGCAACAATAATCTCCCATCCGGGTTT	733
QY	1401	CATTAACACTAGAGAGAACACGCAATGCTAATAAAAAACAGATGTGAGGAGAGAGCCA	1460
DB	734	CATTAACCCCCAAGAAAAACAGCCATGGGATTAATAAACCTTACCTTCGGGAGACAGAAACC	793
QY	1461	C 1461	
DB	794	C 794	
RESULT 7			
LOCUS	B0674807	668 bp	mRNA linear EST 15-JUL-2002
DEFINITION	AGENCOURT_8035172 NIH_MGC_102 Homo sapiens CDNA clone IMAGE:6212672		
ACCESSION	B0674807		
VERSION	B0674807.1	GI:21785641	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 668)		
TITLE	NIH-MGC http://mgc.ncl.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LINC2377 row: c column: 09 High quality sequence stop: 656. Location/Qualifiers 1. 668 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6212672" /clone_lib="NIH_MGC_102" /tissue_type="epidermoid carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: salivary gland; Vector: pOMB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
FEATURES	source		
BASE COUNT	216 a	148 c	167 g 137 t
ORIGIN			

Query Match 31.0%; Score 525.4; DB 14; Length 668;
 Best Local Similarity 99.8%; Pred. No. 4.9e-118;
 Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1169 TCCCTGGGGTACCAGGACGCGCCCAAGGCCCCCGGCTCTCTGGCCCATCG 1228
 DB 1 TCCCTGGGGTACCAGGACGCGCCCAAGGCCCCCGGCTCTCTGGCCCATCG 60

OY 1229 GAGCGGTGGTGGCCCTGGCCCTGCAAAATGAGCCAAACCCGCGACCGAGGACAAATGGCT 1288
 DB 61 GAGCGGTGGTGGCCCTGGCCCTGCAAAATGAGCCAAACCCGCGACCGAGGACAAATGGCT 120

OY 1289 GCGCCGCTCCTGCTGGAAGAACTTACAGACAAATGCTACTATTTTCAGTTGAGAAAGAA 1348
 DB 121 GCGCCGCTCCTGCTGGAAGAACTTACAGACAAATGCTACTATTTTCAGTTGAGAAAGAA 180

OY 1349 TTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAGCTTCAATCTTGTTCATTAACA 1408
 DB 181 TTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAGCTTCAATCTTGTTCATTAACA 240

OY 1409 CTAGAGAGACAGCAATGGATTAATAAACAATGATGGTAGGAGAGAGACCACTGGATCG 1468
 DB 241 CTAGAGAGACAGCAATGGATTAATAAACAATGATGGTAGGAGAGAGACCACTGGATCG 300

OY 1469 GCCTCAGACACTCAGAGCGGTGAATAATGATGAGTGGCTGATGGACATCTCCAGACT 1528
 DB 301 GCCTCAGACACTCAGAGCGGTGAATAATGATGAGTGGCTGATGGACATCTCCAGACT 360

OY 1529 ACAAATTTGGAAGCTGAGACAGCCGATTAAGTGGGCTCATGGCCATGGCCAGAGAGAAG 1588
 DB 361 ACAAATTTGGAAGCTGAGACAGCCGATTAAGTGGGCTCATGGCCATGGCCAGAGAGAAG 420

OY 1589 ACTGAGTGGGCTGATTAATGCTGGGAGAGTGAAGATTTCAATCTGAAGACGTCATAA 1648
 DB 421 ACTGAGTGGGCTGATTAATGCTGGGAGAGTGAAGATTTCAATCTGAAGACGTCATAA 480

OY 1649 ACTTATTTTGGCAAAAGACAGAGAGACAGTACTGTCATCTGCATTA 1695
 DB 481 ACTTATTTTGGCAAAAGACAGAGAGACAGTACTGTCATCTGCATTA 527

RESULT 8
 BE805750/c 511 bp mRNA linear EST 12-JAN-2001
 LOCUS OVI-C10173-071100-464-c02 C10173 Homo sapiens CDNA, mRNA sequence.
 DEFINITION BE805750
 ACCESSION BE805750.1 GI:12134739
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 511)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsushima, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=OVI16t2=OVI-C10173-071100-464-c02&t3=2000-11-07&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 492.
 Location/Qualifiers
 1..511
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="C10173"
 /dev_stage="Adult"
 /note="Organ: colon; lins: Vector: puc18; Site:1: SmaI;
 Site:2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 94 a 113 c 139 g 165 t
 ORIGIN

Query Match 28.8%; Score 488.6; DB 12; Length 511;
 Best Local Similarity 98.4%; Pred. No. 4.7e-109;
 Matches 505; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

OY 27 CGTGCTCAGGCGCAATCTGCAGAACCAATGTAATTCATTAATGTGTCATGAACT 86
 DB 511 CGTGCTCAGGCGCAATCTGCAGAACCAATGTAATTCATTAATGTGTCATGAACT 452

OY 87 CAACAACCTGAACCTGACCCAGGTGAGAGAGAAAGCTCATACGAATCTGCAGCGGT 146
 DB 451 CAACAACCTGAACCTGACCCAGGTGAGAGAGAAAGCTCATACGAATCTGCAGCGGT 392

OY 147 TGTGATGACACAAGCGAGGCTATCCAGGATTCAGAAAGCACTTTCAAAATCTGCAGCA 206
 DB 391 TGTGATGACACAAGCGAGGCTATCCAGGATTCAGAAAGCACTTTCAAAATCTGCAGCA 332

OY 207 GGTATTTCTTCAAGCCAAAGAGACAGGATTTGGCTGAAGAGAAAGTGCAGAGCTTGA 266
 DB 331 GGTATTTCTTCAAGCCAAAGAGACAGGATTTGGCTGAAGAGAAAGTGCAGAGCTTGA 272

OY 267 GAGCGTGGCTGCAACACTGTGCTTTGGCCAAAGCAACAGACACCTGGAGATAT 326
 DB 271 GAGCGTGGCTGCAACACTGTGCTTTGGCCAAAGCAACAGACACCTGGAGATAT 212

OY 327 GAACAGCCAGCTCACTATTTCACAGGTGATGGAAGACATCCACATATCTCTCAAGC 386
 DB 211 GAACAGCCAGCTCACTATTTCACAGGTGATGGAAGACATCCACATATCTCTCAAGC 152

OY 387 CAACGAGCAACCTGAAGACCTGCGAGGACTTACACAAGATGCGAGAAATGAACAGC 446
 DB 151 CAACGAGCAACCTGAAGACCTGCGAGGACTTACACAAGATGCGAGAAATGAACAGC 92

OY 447 CATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGATATTGTGAACAT 506
 DB 91 CATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGATATTGTGAACAT 32

OY 507 CATTTGCAATATCATGTTACACAGCCACCAACCT 539
 DB 31 CATTTGCA--TACGTACACAGCCACCAACCT 3

RESULT 9
 B0127513 581 bp mRNA linear EST 19-APR-2002
 LOCUS B0127513
 DEFINITION B0127513
 1160h06.y1 Melton Normalized Mixed Mouse Pancreas 1 N1-MWS1 Mus
 musculus cDNA clone IMAGE:5946107 5' similar to TR:090M15 090M15
 CSR2.1; mRNA sequence.
 ACCESSION B0127513
 VERSION B0127513.1 GI:20201424
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 581)
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, R.,
 Lemishka, I., Seacore, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A.,
 Schmitt, A., Theising, B., Riter, E., Ronko, I., Bennett, J., Cardenas,
 M., Gibbons, M., McCann, R., Cole, R., Tsagaris, W., Williams, T.,
 Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Other ESTs: 1160h06.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@hbp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center This clone is
 available royalty-free through LBNL; please contact the IMAGE
 consortium (info@image.llnl.gov) for further information
 Seq primer: -40RP from Gibco
 High quality sequence stop: 446.

FEATURES
 source
 1. 581
 Location/Qualifiers

/organism="Mus musculus"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone_1lb="Melton Normalized Mixed Mouse Pancreas 1
 N1-MMS1"
 /sex="Both for embryonic & newborn, male for adult and
 adult islet"
 /dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
 adult, mixed"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five
 libraries representing E10.5/12.5 pancreatic bud, E16.5
 pancreas, newborn pancreas, adult pancreas, and adult
 islets of langerhans were separately constructed using
 Superscript Plasmid Library kit (Life Technologies). cDNA
 was made by oligo-dT priming and size-selected by column
 fractionation. Libraries were amplified once on solid
 support and plasmid DNA from each library was prepared
 and mixed in equal amounts. The mixed library DNA was
 normalized by method #4 from Bonaldo, Lennon, and Soares
 1996 Genome Research 6:791-806; 0.5 microgram
 single-stranded mixed library plasmid DNA was mixed with
 5 micrograms PCR product representing mixed library
 inserts and hybridized to an EcoT of 6. Single-stranded
 (unhybridized) plasmids were isolated by hydroxyapatite
 chromatography and used to make this library."

BASE COUNT 189 a 159 c 114 g 119 t
 ORIGIN

Query Match 28.3%; Score 480.2; DB 14; Length 581;
 Best Local Similarity 89.2%; Pred. No. 5.7e-107;
 Matches 518; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 86 TCACACACCTGACCTGACCCAGTGCAGAGAGAACTCATCAGCAATCTGCAGCGT 145
 DB 1 TCACACACCTGACCTGACCCAGTGCAGAGAGAACTCATCAGCAATCTGCAGCGT 60
 QY 146 CTGTGATGACACAGCAGGCTATCCAGCAATCAAGAACGACTTTCAAAATCTGCAGC 205
 DB 61 CTGTGATGACACAGCAGGCTATCCAGCAATCAAGAACGACTTTCAAAATCTGCAGC 120
 QY 206 AGGTTTCTTCAAGCACAAGACAGCAGATGGCTGGAAGGAAGACGACAGCTGC 265
 DB 121 AGGTTTCTTCAAGCACAAGACAGCAGATGGCTGGAAGGAAGACGACAGCTGC 180

QY 266 AGACGCTGCTGCCAACACTCTGCGTTGGCCAAAGCCAAACAGACACCCTGAGAGATA 325
 DB 181 AGACATTGCGCTGCCAACACTCTGCGTTGGCCAAAGCCAAACAGACACCCTGAGAGATA 240
 QY 326 TGAACGACCGCTCACTCTGTCAGAGTGTGAGGAACATCACCACATCTGTCAG 385
 DB 241 TGAATGACCGCTCACTCTGTCAGAGTGTGAGGAACATCACCACATCTGTCAG 300
 QY 386 CCAACGACGAACTGAAAGACCTGAGGAGCTTTACACAAAGATGAGAGATAAGACAG 445
 DB 301 CCAACGACGAGACCTGAAAGACCTTCAAGAGCTTTACACAAAGATGAGAGATAAGACAG 360
 QY 446 CCAATCAAGTTCAACCAACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTGACA 505
 DB 361 CTGTCAAGTTCAACCAACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTGACA 420
 QY 506 TCATTGCAATATCAGTTACACAGCCACACCTGGGAGCGCTGACAGCAATCTTAATG 565
 DB 421 TCATTGCAATATCAGTTACACAGCCACACCTGGGAGCGCTGACAGCAATCTTAATG 480
 QY 566 AAGTCAGACCACTTGCACAGATACCTTTACCAACACACAGATGATCTGACCTCTGA 625
 DB 481 ATGTTAGACCACTTGCACAGATACCTTTACCAACACACAGATGATCTGACCTCTGA 540
 QY 626 ATTAATACCTGCGCCACATCCGTTTGATTCGTTCTCTC 666
 DB 541 ATTAACACCTGCGCCACATCCGTTTGATTCGTTCTCTC 581

RESULT 10
 AL568743/c 967 bp mRNA 1linear EST 16-FEB-2001
 LOCUS AL568743 L1L_FL002_P11 Homo sapiens cDNA clone CSODE005YH04 3 prime
 DEFINITION
 ' mRNA sequence.
 ACCESSION AL568743
 VERSION AL568743.1 GI:12923387
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 967)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1. 967
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSODE005YH04"
 /clone_1lb="L1L_FL002_P11"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(47) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
 ORIGIN
 Query Match 28.0%; Score 475.4; DB 9; Length 967;
 Best Local Similarity 96.7%; Pred. No. 1.1e-105;
 Matches 504; Conservative 3; Mismatches 9; Indels 5; Gaps 2;

QY 1175 GGGTACGAGCATGCCAGGCCCCCGCCGCTCTGTCCTCAGGAGCGG 1234
|||||
Db 898 GGGTACGAGCATGCCAGGCCCCCGCCGCTCTCAGGAGCGG 843
QY 1235 TGGTCCCCCTGGCCCTGCAAGATGAGCCACCCCGGACCGGAGACAATGGCTCCCGC 1294
|||||
Db 842 TTTGCCCTGSCCTCGCAGATGAGCCACCCCGGACCGGAGACAATGGCTCCCGC 783
QY 1295 CTCACGTGAGAGACTTCACAGCAATGCTACTATTCTTCTAGTGAAGAAATTTTG 1354
|||||
Db 782 CTCACGTGAGAGACTTCACAGCAATGCTACTATTCTTCTAGTGAAGAAATTTTG 723
QY 1355 AGGATGCAAGCTTTCTGTGTGAGACAAGCTTTCACATCTTGTCTTAACACTAGAG 1414
|||||
Db 722 AGGATGCAAGCTTTCTGTGTGAGACAAGCTTTCACATCTTGTCTTAACACTAGAG 663
QY 1415 AGGAACACCAATGATTAATAAACAAGATGTAGGAGAGAGAGCCATGGATCGGCTTCA 1474
|||||
Db 662 AGGAACACCAATGATTAATAAACAAGATGTAGGAGAGAGAGCCATGGATCGGCTTCA 603
QY 1475 CAGACTCAGAGCGTGAATAATGAATGGCTGGATGGGACATCTCCAGACTACAAA 1534
|||||
Db 602 CAGACTCAGAGCGTGAATAATGAATGGCTGGATGGGACATCTCCAGACTACAAA 543
QY 1535 ATTGGAAGCTGGACAGCCGGATTAAGTGGGTCTGATGGCCATGGCCAGAGAGAGACTGTG 1594
|||||
Db 542 ATTGGAAGCTGGACAGCCGGATTAAGTGGGTCTGATGGCCATGGCCAGAGAGAGACTGTG 484
QY 1555 CTGGCTTATTTATCTGTGGGAGCTGGAAAGATTTCCAAATGTGAGAGAGCTCAATTA 1654
|||||
Db 483 CTGGCTTATTTATCTGTGGGAGCTGGAAAGATTTCCAAATGTGAGAGAGCTCAATTA 424
QY 1655 TTTGGCAAAAAGACAGGAGACAGTACTGTCTGTCATTA 1695
|||||
Db 423 TTTGGCAAAAAGACAGGAGACAGTACTGTCTGTCATTA 383

RESULT 11
BE910803 580 bp mRNA linear EST 29-SEP-2000
LOCUS BE910803
DEFINITION 601661855F1 NCI-CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962292 5',
mRNA sequence.
ACCESSION BE910803
VERSION BE910803.1 GI:10407765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 580)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
<http://image.llnl.gov>
Plate: LAM9128 row: m column: 13
High quality sequence stop: 580.
Location/Qualifiers
1..580
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3962292"
/clone_lib="NCI-CGAP_Mam1"
/tissue_type="Tumor, biopsy sample"
/dev_stage="10 months, virgin"

/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPOrt6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 159 a 153 c 161 g 107 t
ORIGIN

Query Match 26 3%; Score 445.6; DB 12; Length 580;
Best Local Similarity 85.5%; Pred. No. 1.8e-98;
Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 959 GCGGCAAGAGCTCTAAAGCTCCAGGCCCCCAAGAGCTCCGCTGTTCCCTGGGAAGC 1018
|||||
Db 1 CGAGCAAGAGCTCTAAAGCTCCAGGCCCCCAAGAGCTCCGCTGTTCCCTGGGAAGC 60
QY 1019 CCGGCCCCCTCAGGCCCCCAGTGGGAGCCAGGCCCCCGGCCCCCAGGCAAGAGAGGAC 1078
|||||
Db 61 CTGGCCCCCTCAGGCCCCCAGTGGGAGCCAGGCCCCCAGGCTCCAGGCAAGAGAGGAC 120
QY 1079 TCCCGGCCCCCTCAGGCCCCCTCGGCTTCCAGGAGCTTCCAGGCAAGCTTGGGGAGCCTG 1138
|||||
Db 121 TCCCGGCCCCCTCAGGCCCCCTCGGCTTCCAGGAGCTTCCAGGCAAGCTTGGGTGAGCCTG 180
QY 1139 GGGTCTGAGACTGCTGGGAGCTGCGAGCTTCTGCGGCTACAGGACATGCCAGGCCCA 1198
|||||
Db 181 GAGTACCTGAGACTGCTGGGAGCTGCGAGCTTCTGCGGCTACAGGAGCTGCCAGGATGCC 240
QY 1199 AGGGCCCCCCCCGCTCTCTGCGCCATCAGAGGCGGTGTGCTCCCTGGCCCTGCAAGATG 1258
|||||
Db 241 AGGGACACACTGCGCCCTCAGGCCCCCTCAGGAGCAATGAGGCAATGGCTGTGCAAGATG 300
QY 1259 AGCCAAACCCCGGACCGGAGGAGCAATGGCTGCCGCCCTCAGTGAAGAACTTCACAGACA 1318
|||||
Db 301 AACCAACCCCGGACCGGAGGAGTGAAGGATGCGGCTCTCAGTGAAGAACTTCACAGATA 360
QY 1319 AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCCAAAGCTTTCTGTGAAG 1378
|||||
Db 361 AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCCAAAGCTTTCTGTGAAG 420
QY 1379 ACAAGCTTTCACATCTGTTTTCATTAACACTAGAGAGAGACAGCAATGATTAATAAAC 1438
|||||
Db 421 ACAAGCTTTCACATCTGTTTTCATTAACACTAGAGAGAGACAGCAATGATTAATAAAC 480
QY 1439 AGATGTAGGAGAGAGAGAGCCACTGATGGCTCCACAGACTCAGAGCGTGAATGAAT 1498
|||||
Db 481 ATACCGTGGGAGAGAGAGAGCCACTGATGGCTCCACAGACTCAGAGCGTGAATGAAT 540
QY 1499 GGAAGTGGCTGATGGAGACATCTCCAGACTACAAAATTG 1538
|||||
Db 541 GGAAGTGGCTGATGGAGAGGCTGATGATTAACAAAACATG 580

RESULT 12
AW958053 552 bp mRNA linear EST 01-JUN-2000
LOCUS AW958053
DEFINITION ES3730123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW958053
VERSION AW958053.1 GI:8147736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharep, S., Gaspar, R., Gay, C., Holt
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research

ACCESSION BF856923
 VERSION BF856923.1 GI:12244667
 KEYWORDS EST
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 410)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W.J., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=0V1&c2=0V1-FT0201-021100-453-e01&t3=2000-11-02&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 410.
 Location/Qualifiers
 1..410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FT0201"
 /dev_stage="Adult"
 /note="Organ: prostate tumor; Vector: puc18; Site:1; SmaI; Site:2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 73 a 93 c 109 g 135 t
 ORIGIN
 Query Match 23.4%; Score 395.8; DB 12; Length 410;
 Best Local Similarity 99.3%; Pred. No. 2.5e-86;
 Matches 408; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
 Oy 130 AGGATCTGACGGCTCTGTGATGACACAGCCAGGCTATCCAGGAATCAAGAACGAC 189
 Db 410 AGGAATCTGACGGCTCTGTGATGACACAGCCAGGCTATCCAGGAATCAAGAACGAC 351
 Oy 190 TTTCAAATCTGACGAGGTTTCTTCAAGCCAAAGAGACAGGATGGCGAAGGAG 249
 Db 350 TTTCAAATCTGACGAGGTTTCTTCAAGCCAAAGAGACAGGATGGCGAAGGAG 291
 Oy 250 AAAGTGCAGAGCTTGACAGCGTGGCTGCCAACACTGTGCTGGCCAAAGCCAAAC 309
 Db 290 AAAGTGCAGAGCTTGACAGCGTGGCTGCCAACACTGTGCTGGCCAAAGCCAAAC 231
 Oy 310 GACACCTTGAGAGATGTAACAGCCAGCTCAATCTATTCAAGGTGATGAGTAAGAACATC 369
 Db 230 GACACCTTGAGAGATGTAACAGCCAGCTCAATCTATTCAAGGTGATGAGTAAGAACATC 171
 Oy 370 ACCACTATCTCAAGCCAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 429
 Db 170 ACCACTATCTCAAGCCAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGT 111
 Oy 430 GCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAAGCTTCCAGCTCTTTGAG 489

Db 110 GCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAAGCTTCCAGCTCTTTGAG 51
 Oy 490 ACCGATATTGTGACATCATTCATATTCAGTTACACAGCCACCACCTG 540
 Db 50 ACCGATATTGTGACATCATTCATATTCAGTTACACAGCCACCACCTG 1
 RESULT 15
 LOCUS BB248064 638 bp mRNA linear EST 23-OCT-2001
 DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730023E20 3', mRNA sequence.
 ACCESSION BB248064
 VERSION BB248064.2 GI:16355610
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 638)
 AUTHORS Arikawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arikawa, T., et al. 2001)
 Unpublished (2001)
 On Jul 6, 2000 this sequence version replaced gi:8940810.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1571-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Location/Qualifiers
 1..638
 /organism="Mus musculus"
 /db_xref="taxon:10090"

BASE COUNT	212 a	134 c	146 g	146 t.
ORIGIN	/clone="A730023E20" /clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum" /tissue_type="cerebellum" /dev_stage="7 days neonate" /lab_host="DH10B" /note="Site.1: SalI; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGAGAGAGAGATCCAGAGGCTCTTTTCTTTTCTTTVN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATTCGAGGTTAAATTAATCCCCCCCCCC 3'. cDNA was cleaved with XhoI and BamHI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I."			

Query Match	21.8%	Score 369.4	DB 10	Length 638
Best Local Similarity	84.5%	Pred. No. 9.4e-80		
Matches 415	Conservative 0	Mismatches 76	Indels 0	Gaps 0

Oy	1205	CCCCGGCCCTCTGGGCCCTTCAAGAGAGCGGTGGTGGCCCTGGCCCTGGAGATGAGCCA	1264
Db	1	CACCTGGCCCTCCAGGCCCTCTCAGAGAGCAATGGAGCCATTGGCTCTGCAGATGACCAA	60
Oy	1265	CCCCGGCACCGGAGAGACATGCTGCCGCCCTCTACTGGAAGACTTTCACAGACAATGCT	1324
Db	61	CCCCGAGCTCAGAGGTCACAGCGATGTCCGCTCACAGGAAGAACTTTCACAGATTAATGCT	120
Oy	1325	ACTATTTTTCAGTGTAGAGAAATTTTGGAGATGCCAAGCTTTCTGTGTAAGACAGT	1384
Db	121	ACTATTTTTCATGTGAAAAAGAAATTTTGGAGATGCTTAAGCTTTTCTGTGAAGACAAT	180
Oy	1385	CTTCCACATCTTGTTTTCATTAACACTAGAGAGGAACAGCAATGATTAAAAAACAGATGG	1444
Db	181	CTTCCCATCTCGTTTTCATTAACCTCAAGAGAGAAACAGCAATGATTAAAAAAGCATACCG	240
Oy	1445	TAGGGAGAGAGACCCACTGGATGGCCCTCCACAGACTCCAGCGTAAATAATGAATGAGT	1504
Db	241	TGGGGAGAGAAACCCATTGGATGGCTCCACAGACTCCAAACAGGAAAGCGAATGGAAGT	300
Oy	1505	GGCTGATGTGGGACATCTCCAGACTACAAAAATTGGAAAGCTGAGAGCCGGATTAACCTGG	1564
Db	301	GGCTAGACGGGGTACACCTGTGATTACAAAACTGGAAACCTGGACACCAAGATTAACCTGG	360
Oy	1565	GTCAATGGCCATGGGCCAGAGAGAACTGTGCTGGTGAATTATCTGGCCAGTGGAAAGC	1624
Db	361	GCAATGGCCATGGGCCAGAGAGAAAGTGTGCTGGTGAATTATCCAGAGACAGTGAATG	420
Oy	1625	ATTTTCCATGTGAAGAGCTCAATTAACCTTAATTTGCGAAAAAAGACAGAGACAGTACTGT	1684
Db	421	ACTTCCAGTGTGAATCAATTAACCTTAATTTGTGSAAGGAAGGAAGGAGCGATACAT	480
Oy	1685	CATCTGCATTA 1695	
Db	481	CATCCATATTA 491	

Search completed: March 21, 2003, 08:07:57
Job time : 2240.7 secs

TITLE The membrane-type collectin CL-P1 is a scavenger receptor on
vascular endothelial cells
JOURNAL J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE 21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani, K.
TITLE Direct Submission
Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1 Midori-igaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp,
Tel: +81-166-68-2393, Fax: +81-166-68-2399)
FEATURES
Location/Qualifiers
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue.lib="placenta"
1..2983
/gene="CL-P1"
71..2299
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDFAEEEEVVSFGYKRFQIEGTQCTCKNNMALKESIIILY
ILCALITIVAILGYKVEKMDVVTGMEISRQYDQDKLTAVESDKLADQTKKAI
STNSELSTERSLIDLRQOLREITERTSKNKLDELQASGDLVROSLQKTELENN
SFLITVNNKTLQANGVYNNLODTSVLQGNLQNNYSHNVVIMNNLNINLQVOORN
LITNLORSVDNLSOALFTQOMENITTSQANEONLKDLDLKHDAENRIAIKNOLEE
RQLEFEDIVNIIISNISTVAHLRLTSLNENRTCTDPLTKHBDLTSLNMTLANI
RLDSVSLRMQODLMRSRLDTEVANLIVIMEMLVDSKHQILKNPTILOGPGRGP
GRSGRSGQPGPPTGNKGKGEKEPPEPAGREPIGAPPEGEKSGSKSQPK
GSGRSGKPGQSGDPGPGPGPKGELGEPGPGQSGLOGTPEGPGYGRGLG
LPGPGMPGKPGPBPSPGSAVPLALNEPAPEDNCCGPHMKNEPDKCYEYVE
KEIFEDAKLFCEDKSSHYLFINTREOQMKKQMGVRESHWICLTSERENEMKMDG
TSPDYNNMKAQPDNMNGHGHGPEDCAGLTYAGWMDPCEDEVNPFICEKRETVLSS
AL"

BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN

Query Match 100.0%; Score 1371; DB 9; Length 2983;
Best Local Similarity 100.0%; Pred. No. 2.2e-308;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACAGCAGCTCACTCATTACAGTCAGATGAGAAATCACCACATCTCTCAA 60
|||||
Db 926 ATGAACAGCAGCTCACTCATTACAGTCAGATGAGAAATCACCACATCTCTCAA 985
QY 61 GCCAAGCAGCAGAACTGAAAGACCTTGACAGACTTACACAAAGATGAGAAATGAGACA 120
|||||
Db 986 GCCAAGCAGCAGAACTGAAAGACCTTGACAGACTTACACAAAGATGAGAAATGAGACA 1045
QY 121 GCCATCAAGTTCAACCACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTGAAC 180
|||||
Db 1046 GCCATCAAGTTCAACCACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTGAAC 1105
QY 181 ATCATAGCAATATCAGTTACACAGCCACACCTGCGAGCGCTGACAGCAATCTAAAT 240
|||||
Db 1106 ATCATAGCAATATCAGTTACACAGCCACACCTGCGAGCGCTGACAGCAATCTAAAT 1165
QY 241 GAAGTCAGAGCACTTGACAGATACCTTACCAACACACAGATGATGACCTCCTTG 300
|||||
Db 1166 GAAGTCAGAGCACTTGACAGATACCTTACCAACACACAGATGATGACCTCCTTG 1225
QY 301 AATTAATACCTGGCAACATCCGTTGGATTCTGTTCTCTCAGATGACAAAGATTGG 360
|||||
Db 1226 AATTAATACCTGGCAACATCCGTTGGATTCTGTTCTCTCAGATGACAAAGATTGG 1285
QY 361 ATGAGGTCGAGGTTAGACACTGAAGTGCACACTATCATGATGAGAAAGAAATGAG 420
|||||
Db 1286 ATGAGGTCGAGGTTAGACACTGAAGTGCACACTATCATGATGAGAAAGAAATGAG 1345

QY 421 CTAGTAGACTCCAAAGCATGTGCTCAGCTCATCAAGAAATTTTACAATACTACAGGTCACCG 480
|||||
Db 1346 CTAGTAGACTCCAAAGCATGTGCTCAGCTCATCAAGAAATTTTACAATACTACAGGTCACCG 1405
QY 481 GGGCCCAAGGGGTCCAGAGGTGACAGAGATCCAGAGGACCCCTGGCCCACTGGCAAC 540
|||||
Db 1406 GGGCCCAAGGGGTCCAAAGGTGACAGAGATCCAGAGGACCCCTGGCCCACTGGCCAAAC 1465
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGACCTGGACCACTGGCCCTGGGGTGGAGAGGC 600
|||||
Db 1466 AAGGACAGAAAGAGAGAGAGGGGAGACCTGGACCACTGGCCCTGGGGTGGAGAGGC 1525
QY 601 CCAATTGGACCACTGTGCTCCCGCGAGAGCGTGGCGGCAAGAGATCAAGGCTCCACG 660
|||||
Db 1526 CCAATTGGACCACTGTGCTCCCGCGAGAGCGTGGCGGCAAGAGATCAAGGCTCCACG 1585
QY 661 GGGCCCAAGGCTCCGCTGCTTCCCTGGGAGAGCCCGGCTTCAAGGCCCCCACTGGGGAC 720
|||||
Db 1586 GGGCCCAAGGCTCCGCTGCTTCCCTGGGAGAGCCCGGCTTCAAGGCCCCCACTGGGGAC 1645
QY 721 CCAAGGCCCCCGGGCCACCAAGGCAAGAGAGGACTCCCGGCTTCAAGGCCCCCACTGGGG 780
|||||
Db 1646 CCAAGGCCCCCGGGCCACCAAGGCAAGAGAGGACTCCCGGCTTCAAGGCCCCCACTGGGG 1705
QY 781 TTCAGGAGACTTGAAGGACCGCTGGGGAGCGTGGGTGCTTGAGACTCGGGGACTGCCA 840
|||||
Db 1706 TTCAGGAGACTTGAAGGACCGCTGGGGAGCGTGGGTGCTTGAGACTCGGGGACTGCCA 1765
QY 841 GCGTGGCTGGGGTACAGGACGATGCCAGGGCCCCCGGCCCCCTCTGGGCCCA 900
|||||
Db 1766 GCGTGGCTGGGGTACAGGACGATGCCAGGGCCCCCGGCCCCCTCTGGGCCCA 1825
QY 901 TCAAGAGCGGTGGTGGCCCCCTGGCCCTGCAAAATGAGCAACCCCGGACCGAGAGCAAT 960
|||||
Db 1826 TCAAGAGCGGTGGTGGCCCCCTGGCCCTGCAAAATGAGCAACCCCGGACCGAGAGCAAT 1885
QY 961 GCGTGGCCCCCTGACCTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
|||||
Db 1886 GCGTGGCCCCCTGACCTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1945
QY 1021 GAAATTTTGAAGATGGAAGCTTTTCTGTAAGACAGCTTCACATCTGTTTTCATA 1080
|||||
Db 1946 GAAATTTTGAAGATGGAAGCTTTTCTGTAAGACAGCTTCACATCTGTTTTCATA 2005
QY 1081 AACACTAGAGAGAGACAGCAATGGATTAATAAACAAGATGTTAGAGAGAGAGACCACTGG 1140
|||||
Db 2006 AACACTAGAGAGAGAGCAATGGATTAATAAACAAGATGTTAGAGAGAGAGACCACTGG 2065
QY 1141 ATGGGCTTCACAGACTGAGAGCGTGAATGAATGAATGAGTGGCTGGATGGACATCTCCA 1200
|||||
Db 2066 ATGGGCTTCACAGACTGAGAGCGTGAATGAATGAATGAGTGGCTGGATGGACATCTCCA 2125
QY 1201 GACTACAAAAAATGGAAGGTGACAGCGGATTAACCTGGGGTATGGCCATGGGCCACAGA 1260
|||||
Db 2126 GACTACAAAAAATGGAAGGTGACAGCGGATTAACCTGGGGTATGGCCATGGGCCACAGA 2185
QY 1261 GAAGACTGTGCTGGTGTGATTTATGCTGGGACGTGAACATTTCCAAATGTAAGACGTC 1320
|||||
Db 2186 GAAGACTGTGCTGGTGTGATTTATGCTGGGACGTGAACATTTCCAAATGTAAGACGTC 2245
QY 1321 AATTAATCTATTTGCGAAAAAGACAGGAGACAGTACTGTCAATCTGCATTA 1371
|||||
Db 2246 AATTAATCTATTTGCGAAAAAGACAGGAGACAGTACTGTCAATCTGCATTA 2296

RESULT 2
AX047353 2641 bp DNA linear PAT 15-DEC-2000
LOCUS AX047353
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 2641)
 AUTHORS Bandman, O., Hillman, J.L., Tang, Y.T., Lai, P., Yue, H., Baughn, M.R.,
 Lu, D.A. and Azimzal, Y.
 TITLE Extracellular matrix and adhesion-associated proteins
 JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
 Incyte Genomics, Inc. (US)
 FEATURES
 source location/Qualifiers
 1..2641
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 4586187CBI"
 BASE COUNT 823 a 628 c 608 g 582 t
 ORIGIN
 Query Match 99.9%; Score 1369.4; DB 6; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 5.2e-308;
 Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 0;
 QY 1 ATGACAGCCAGCTCACTCATTCACAGGTGAGAGAACATCCACTATCTCTCAA 60
 |||||
 Db 567 ATGACAGCCAGCTCACTCATTCACAGGTGAGAGAACATCCACTATCTCTCAA 626
 QY 61 GGCACGAGCAGACCTGAAAGCCTGACGAGCTTACAAAGATGACAGATAGAAC 120
 |||||
 Db 627 GGCACGAGCAGACCTGAAAGCCTGACGAGCTTACAAAGATGACAGATAGAAC 686
 QY 121 GCCATCAAGTTCAACCACTGGAGAACGTTCCAGCTCTTGGAGAGATATTGGAAC 180
 |||||
 Db 687 GCCATCAAGTTCAACCACTGGAGAACGTTCCAGCTCTTGGAGAGATATTGGAAC 746
 QY 181 ATCATTAGCAATATCATGTTACACAGCGCCACCACTCGGAGCGCTGACCAATCTAAT 240
 |||||
 Db 747 ATCATTAGCAATATCATGTTACACAGCGCCACCACTCGGAGCGCTGACCAATCTAAT 806
 QY 241 GAAGTCAGGACCACTGACAGATACCTTACCAACAGATATGACCTGCTTGG 300
 |||||
 Db 807 GAAGTCAGGACCACTGACAGATACCTTACCAACAGATATGACCTGCTTGG 866
 QY 301 AATTAATACCTGCGCCAACTCCGTTTGATTTCTCTCTAGATGCAACAGATTGG 360
 |||||
 Db 867 AATTAATACCTGCGCCAACTCCGTTTGATTTCTCTCTAGATGCAACAGATTGG 926
 QY 361 ATGAGGTGCAAGTTTACACTGATAGCCCAATTAATGATGATATGGAAGATGAG 420
 |||||
 Db 927 ATGAGGTGCAAGTTTACACTGATAGCCCAATTAATGATGATATGGAAGATGAG 986
 QY 421 CTAGTAGACTCCAGCATGCTGATCATCAGAAATTTTACAAATTTACAGGTCCACCG 480
 |||||
 Db 987 CTAGTAGACTCCAGCATGCTGATCATCAGAAATTTTACAAATTTTACAGGTCCACCG 1046
 QY 481 GGGCCAGGAGGTCACAGAGGTGACAGAGATCCAGAGGACCCCTGGCCCAATGCGCAAC 540
 |||||
 Db 1047 GGGCCAGGAGGTCACAGAGGTGACAGAGATCCAGAGGACCCCTGGCCCAATGCGCAAC 1106
 QY 541 AAGGGACAAAGAGAGAGAGAGGGGAGCTTGACACACTGCGCCCTCGGGGTGAGAGAGC 600
 |||||
 Db 1107 AAGGGACAAAGAGAGAGAGAGGGGAGCTTGACACACTGCGCCCTCGGGGTGAGAGAGC 1166
 QY 601 CCAATTGGACCAAGTGTGCTCCCGGAGAGCGTGGGGAAGAGATCTAAAGGCTCCAG 660
 |||||
 Db 1167 CCAATTGGACCAAGTGTGCTCCCGGAGAGCGTGGGGAAGAGATCTAAAGGCTCCAG 1226
 QY 661 GGGCCCAAGAGGTCCTGCTGTTCCCTCGGGAAGCCCGGCTTCAGGAGCCCAATGCGGAG 720
 |||||
 Db 1227 GGGCCCAAGAGGTCCTGCTGTTCCCTCGGGAAGCCCGGCTTCAGGAGCCCAATGCGGAG 1286
 QY 721 CCAAGGCCCCCGGCGCCACAGGCAAGAGAGACTCCCGGCGCTCAGAGGCGCTCTGCGC 780
 |||||
 Db 1287 CCAAGGCCCCCGGCGCCACAGGCAAGAGAGACTCCCGGCGCTCAGAGGCGCTCTGCGC 1346

QY 781 TTCACGAGACTTCAGGGACCCGTTGGGAGACCTGGGGGTGCTGACCTGGGACTGCCA 840
 |||||
 Db 1347 TTCACGAGACTTCAGGGACCCGTTGGGAGACCTGGGGGTGCTGACCTGGGACTGCCA 1406
 QY 841 GGGTTCCTGGGGTACACGAGCATGCCAGGCCCGCCCGGCGCTCTGCGGCCA 900
 |||||
 Db 1407 GGGTTCCTGGGGTACACGAGCATGCCAGGCCCGCCCGGCGCTCTGCGGCCA 1466
 QY 901 TCAGAGCGGTGTGCTGCTGCGCTGCGAGATAGACCAACCCCGGCGGAGAGACAT 960
 |||||
 Db 1467 TCAGAGCGGTGTGCTGCTGCGCTGCGAGATAGACCAACCCCGGCGGAGAGACAT 1526
 QY 961 GGGTGGCCCGCTACTGGAAGAACTTCACAGCAATGCTACTATTTTTCAGTTGAGAA 1020
 |||||
 Db 1527 AGCTGCCCGCTTACTGGAAGAACTTCACAGCAATGCTACTATTTTTCAGTTGAGAA 1586
 QY 1021 GAAATTTTGGAGATCCAAAGCTTTTCTGTGAAGACAGTCTTCACATCTTGTTCATA 1080
 |||||
 Db 1587 GAAATTTTGGAGATCCAAAGCTTTTCTGTGAAGACAGTCTTCACATCTTGTTCATA 1646
 QY 1081 AACCTAGAGAGAGACAGCAATGATTAATAAACAAGATGTTAGGAGAGAGACCACTGG 1140
 |||||
 Db 1647 AACCTAGAGAGAGAGACAGCAATGATTAATAAACAAGATGTTAGGAGAGAGACCACTGG 1706
 QY 1141 ATGGGCTCACAGACACAGCGGTGAATGAATGGAAGCGCTGGATGGACATCTCCA 1200
 |||||
 Db 1707 ATGGGCTCACAGACACAGCGGTGAATGAATGGAAGCGCTGGATGGACATCTCCA 1766
 QY 1201 GACTACAAAATTTGAAAGCTGGACAGCCGATATGAGGAGTATGAGCCATGGGCCAGGA 1260
 |||||
 Db 1767 GACTACAAAATTTGAAAGCTGGACAGCCGATATGAGGAGTATGAGCCATGGGCCAGGA 1826
 QY 1261 GAAGACTGTGCTGGTTGATTTATGCTGGGAGCTGGAAGATTTCCAAATGTGAAGACGTC 1320
 |||||
 Db 1827 GAAGACTGTGCTGGTTGATTTATGCTGGGAGCTGGAAGATTTCCAAATGTGAAGACGTC 1886
 QY 1321 AATTAATCAATTTGGCAAAAAGACAGGAGACGCTGCTGATCTCATCTCAATTA 1371
 |||||
 Db 1887 AATTAATCAATTTGGCAAAAAGACAGGAGACGCTGCTGATCTCATCTCAATTA 1937

RESULT 3
 AB038518
 LOCUS AB038518 3058 bp mRNA linear PRI 08-MAR-2001
 DEFINITION Homo sapiens SCRL mRNA for scavenger receptor with C-type lectin
 type 1, complete cds.
 ACCESSION AB038518
 VERSION AB038518.1 GI:13365514
 KEYWORDS
 SOURCE Homo sapiens tissue_l1b:Placenta CDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F., and Nakamura, T.
 TITLE Molecular cloning and functional characterization of a human
 scavenger receptor with C-type lectin (SCRL), a novel member of a
 scavenger receptor family
 JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
 MEDLINE 21092718
 REFERENCE 2 (bases 1 to 3058)
 AUTHORS Nakamura, K. and Nakamura, T.
 TITLE Direct Submission
 JOURNAL Direct Submission
 Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
 School, Division of Biochemistry, Biomedical Research Center, 2-2
 Yamadaoka, Suita, Osaka 565-0871, Japan
 (E-mail: knakamura@onbich.med.osaka-u.ac.jp,
 Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
 FEATURES
 source location/Qualifiers
 1..3058
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

```
gene /tissue_lib="Placenta"
      1..3058
      /gene="SRCL"
      209..2437
      /gene="SRCL"
      /codon_start=1
      /product="scavenger receptor with C-type lectin type I"
      /protein_id="BAB39147.1"
      /db_xref="GI:11365515"
      /translation="MKDDFAEEVEVSFGYKRRFGIOGTCTCKNNMLKFSIILY
      IICALLITVALIGYKVERKMDVNTGSMETSTROTVDKLTAVSDLKIKIGDQKKAI
      STNSELSTFRSDILDLRQDLRETEKTSKNDLEKLOASGDVLRYROSLKTELENN
      SPDLTVNKTTLQNGYVTNLDQDTSVLQGNLQNMVSHVNVIMNINLNLVQOQRN
      L1TNLQSRVDDTSQALQRIKNDQNLQVLAQKQDMLKEVQSLQTLAANNSLA
      KANNDLEDNMSQNSFTQMENTITTSQANEQNLKDLQDLKDAENRTAIKENOLE
      ROLFETDIYINISNTSYAHILRLTLNLEVRTCTDPLDHLKTDLSLNTLANI
      RLDVSILRMQOOLDRSLDREYANLILSVIMEKMLVDSKGLIKNTITIOGPRGPR
      RQDRSQGPRGPTGNKGQKEKEBPSPRACRGIRGIRGAPRGEKGKSKSQPK
      GSRGSGKPGPQPSGDPGPPGPKEGLEPGPGPGGLOGTVEPGVPGRLPG
      LRGVPMGPGRPPGPPGSAVPLALNEPRAEDNSCPHWNFNFDKCYFSE
      KELEFDKLFCEDEKSSHLVEINTREDOQWIKQMVRESHMLGLTDSERENEMKMDI
      TSPDYKNMKAGQPDNMHGHGPEDCAGLILYAGQNMDFQCEDEVNFIICEKDRETVLS
      AL"
polya_signal 3020..3025
            /gene="SRCL"
BASE COUNT 911 a 744 c 764 g 639 t
ORIGIN
Query Match      99.9%; Score 1369.4; DB 9; Length 3058;
Best Local Similarity 99.9%; Pred. No. 5.2e-308;
Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY  * 1 ATGAACAGCAGCTCAACTCATTTCACAGGTTCAGATGAGAAATCACACTATCTCTCAA 60
DB 1064 ATGAACAGCAGCTCAACTCATTTCACAGGTTCAGATGAGAAATCACACTATCTCTCAA 1123
QY 61 GCCACAGAGAGAACTCGAAGAACCTGACAGACTTACACAAAGTGCAGAGATAGACAA 120
DB 1124 GCCACAGAGAGAACTCGAAGAACCTGACAGACTTACACAAAGTGCAGAGATAGACAA 1183
QY 121 GCCATCAAGTTCAACCACTGAGAGAACGCTTCAGCTCTTGGAGAGATATTTGTGAC 180
DB 1184 GCCATCAAGTTCAACCACTGAGAGAACGCTTCAGCTCTTGGAGAGATATTTGTGAC 1243
QY 181 ATGATTAGCAATATCAGATTACACAGCCACACCTGCGAGCGTGACACAGAACTCTAAAT 240
DB 1244 ATGATTAGCAATATCAGATTACACAGCCACACCTGCGAGCGTGACACAGAACTCTAAAT 1303
QY 241 GAAGTGCAGACACTTGCACAGATACCTTACCAAAACACACATGATGATGACTCTCTG 300
DB 1304 GAAGTGCAGACACTTGCACAGATACCTTACCAAAACACACATGATGATGACTCTCTG 1363
QY 301 AATTAATACCTGCGCAACATCCGTTTGATTTCTCTCAGAGATGACAAAGATTTG 360
DB 1364 AATTAATACCTGCGCAACATCCGTTTGATTTCTCTCAGAGATGACAAAGATTTG 1423
QY 361 ATGAGGTGAGAGTTAGACACTGAGTAGCCAACTTATCAGTATGATGAGAAATGAG 420
DB 1424 ATGAGGTGAGAGTTAGACACTGAGTAGCCAACTTATCAGTATGATGAGAAATGAG 1483
QY 421 CTGAGTACATCCCAAGCATGTGTCAGCTCATCAAGAAATTTTACATACAGAGTCCACCG 480
DB 1484 CTGAGTACATCCCAAGCATGTGTCAGCTCATCAAGAAATTTTACATACAGAGTCCACCG 1543
QY 481 GGCCCCCAGGGGTCCAAGAGTGCAGAGAGATCCAGAGGACCCCTGCGCCCAACTGGCAAC 540
DB 1544 GGCCCCCAGGGGTCCAAGAGTGCAGAGAGATCCAGAGGACCCCTGCGCCCAACTGGCAAC 1603
QY 541 AAGGACAGAAAGAGAGAAAGGGGAGACTGAGACACACTGCGCCCTGCGGGTGAAGAGAGC 600
DB 1604 AAGGACAGAAAGAGAGAAAGGGGAGACTGAGACACACTGCGCCCTGCGGGTGAAGAGAGC 1663
QY 601 CCAATTGGACAGCTGTGTCCTCCCGGAGAGCGTGCGCGCAAAAGATCTAAAGGCTCCACG 660
```

```
DB 1664 CCAATTGGACAGCTGTGTCCTCCCGGAGAGCGTGCGCGCAAAAGATCTAAAGGCTCCACG 1723
QY 661 GGCCCCCAAGAGCTCCGCTGGTTCCCTCCCGGGAGAGCGCGCCCTCAGAGGCCCCAGTGGGGAG 720
DB 1724 GGCCCCCAAGAGCTCCGCTGGTTCCCTCCCGGGAGAGCGCGCCCTCAGAGGCCCCAGTGGGGAG 1783
QY 721 CCAAGGCCCCCGGGCCACCAAGAGCAAGAGAGGAGCTCCCGCGCCCTCAGAGGCCCCAGTGGGGAG 780
DB 1784 CCAAGGCCCCCGGGCCACCAAGAGCAAGAGAGGAGCTCCCGCGCCCTCAGAGGCCCCAGTGGGGAG 1843
QY 781 TTCCAGGAGACTTCAAGGACACCGTTGGGAGACCTGGGGTGCCTGGAGCTCGGGAGACTGCCA 840
DB 1844 TTCCAGGAGACTTCAAGGACACCGTTGGGAGACCTGGGGTGCCTGGAGCTCGGGAGACTGCCA 1903
QY 841 GGCTTGCTGGGGTACACAGGATGCGAGGCCCAAGGGCCCCCGCGCCCTCCTGGGCGCA 900
DB 1904 GGCTTGCTGGGGTACACAGGATGCGAGGCCCAAGGGCCCCCGCGCCCTCCTGGGCGCA 1963
QY 901 TCAGGAGCGGTGTGTCCTGCGCCCTGCGAGATGAGCCAAACCCCGGACCGGAGAGCAAT 960
DB 1964 TCAGGAGCGGTGTGTCCTGCGCCCTGCGAGATGAGCCAAACCCCGGACCGGAGAGCAAT 2023
QY 961 GGCTGCGCCGCTCAGCTGAGAGACTTCACAGACAAATGCTATTTTCAGTTGAGAA 1020
DB 2024 AGCTGCGCCGCTCAGCTGAGAGAACTTCACAGACAAATGCTATTTTCAGTTGAGAA 2083
QY 1021 GAAATTTTGGAGTGAAGTGAAGCTTTTCTGTGAAGACAACTTCACATCTTGTTCATAT 1080
DB 2084 GAAATTTTGGAGTGAAGTGAAGCTTTTCTGTGAAGACAACTTCACATCTTGTTCATAT 2143
QY 1081 AACACTGAGAGAGAACAGCAATGATATAAATAACAGATGCTAGGAGAGAGACCACTG 1140
DB 2144 AACACTGAGAGAGAACAGCAATGATATAAATAACAGATGCTAGGAGAGAGACCACTG 2203
QY 1141 ATGGGCTCAGACACTGAGAGCGTGAATAATGAAATGAGAGTGGTGGATGGACATGCCA 1200
DB 2204 ATGGGCTCAGACACTGAGAGCGTGAATAATGAAATGAGAGTGGTGGATGGACATGCCA 2263
QY 1201 GACTACAAAATTTGGAAGCTGAGACAGCCGATTAACCTGGGGTATGGCCATGGGCCAGGA 1260
DB 2264 GACTACAAAATTTGGAAGCTGAGACAGCCGATTAACCTGGGGTATGGCCATGGGCCAGGA 2323
QY 1261 GAAGACTGCTGCTGGGTGATTTATGCTGGCAGTGAAGCAATTTCCAAATGTAAGAGCTG 1320
DB 2324 GAAGACTGCTGCTGGGTGATTTATGCTGGCAGTGAAGCAATTTCCAAATGTAAGAGCTG 2383
QY 1321 AATTACTTCAATTTGCGAAAAGAGAGAGACAGTACTGTCATCTGCATTA 1371
DB 2384 AATTACTTCAATTTGCGAAAAGAGAGAGACAGTACTGTCATCTGCATTA 2434

RESULT 4
AX454442
LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 27 from Patent WO0208284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS
SOURCE
ORGANISM human.
REFERENCE
1 Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Gunney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
Paoni,N.F., Stephan,U.P., Watanabe,C.K., Williams,P.M., Wood,W.I.
and Ye,W.
Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
Patent: WO 0208284-A 27 31-JAN-2002:
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gervlsen, Mary E. (US) ; Goddard,
```

Andrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
 Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
 ; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
 Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
 I. (US)

FEATURES
 source location/Qualifiers
 1..2005
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 606 a 493 c 491 g 415 t
 ORIGIN

Query Match 99.0%; Score 1357.4; DB 6; Length 2005;
 Best Local Similarity 99.9%; Pred. No. 3.2e-305;
 Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 ATGAACAGCCAGCTCACTCAATTCACAGAGTGCAGATGAGACATCAACCACTATCTCTCAA 60
 134 ATGAACAGCCAGCTCACTCAATTCACAGAGTGCAGATGAGACATCAACCACTATCTCTCAA 193
 61 GCCAAGAGGAGACCTGAAAGACCTGCAGAGCTTACACAAAGATGACAGATAGACAA 120
 134 GCCAAGAGGAGACCTGAAAGACCTGCAGAGCTTACACAAAGATGACAGATAGACAA 253
 121 GCCATCAAGTTCAACCACTGAGAGAGCGCTCCAGCTCTTGAGAGCGATATTGTGAAC 180
 254 GCCATCAAGTTCAACCACTGAGAGAGCGCTCCAGCTCTTGAGAGCGATATTGTGAAC 313
 181 ATCATTTAGCAATATACAGTTACACAGCCACACCTGCGAGCCTGCACGCAATCTAAT 240
 314 ATCATTTAGCAATATACAGTTACACAGCCACACCTGCGAGCCTGCACGCAATCTAAT 373
 241 GAAGTCAGAGCACTTGACATATACCTTACCAAAACACAGATGATGTGACCTCTCTG 300
 374 GAAGTCAGAGCACTTGACATATACCTTACCAAAACACAGATGATGTGACCTCTCTG 432
 301 AATAATACCTTGCCCAACATCCGTTTGATTCCTCTCTCAGATGCAAGATTTG 360
 433 AATAATACCTTGCCCAACATCCGTTTGATTCCTCTCTCAGATGCAAGATTTG 492
 361 ATGAGTGTGAGTTAGACACTGAAGTACCACTTATCTAGTATTTGGAAGAAATGAAG 420
 493 ATGAGTGTGAGTTAGACACTGAAGTACCACTTATCTAGTATTTGGAAGAAATGAAG 552
 421 CTAGTAGACTCCAGCATGCTGAGCTCATCAAGATTTTACATATCTACAAGTCCACCG 480
 553 CTAGTAGACTCCAGCATGCTGAGCTCATCAAGATTTTACATATCTACAAGTCCACCG 612
 481 GGCCCCAGGGGTCCAAGAGTACAGAGATCCCAAGAGACCCCTGCGCCCACTGGCAAC 540
 613 GGCCCCAGGGGTCCAAGAGTACAGAGATCCCAAGAGACCCCTGCGCCCACTGGCAAC 672
 541 AAGGACAGAAAGAGAGAGAGGGGAGCTTGACCACTGCGCCCTGCGGTACAGAGCC 600
 673 AAGGACAGAAAGAGAGAGAGGGGAGCTTGACCACTGCGCCCTGCGGTACAGAGAGC 732
 601 CCAATTGGACAGCTGCTGCTCCCGGAGAGCGTGGGGGCAAAAGATCTAAAGCTCCAG 660
 733 CCAATTGGACAGCTGCTGCTCCCGGAGAGCGTGGGGGCAAAAGATCTAAAGCTCCAG 792
 661 GGCCCCAAAGAGCTCCGTTGTTCCCTTGGAAGCCCGGCTTCAGAGGCCCAAGTGGGAC 720
 793 GGCCCCAAAGAGCTCCGTTGTTCCCTTGGAAGCCCGGCTTCAGAGGCCCAAGTGGGAC 852
 721 CCAGGCCCCCGGGCCCAAGAGAGAGGATCTCCCGGCTTCAGAGGCCCAAGTGGGAC 780
 853 CCAGGCCCCCGGGCCCAAGAGAGAGGATCTCCCGGCTTCAGAGGCCCAAGTGGGAC 912
 781 TTCAGAGGACTTCAGAGGACGCTTGGGGAGCTTGGGGTCTCTGAGACCTGAGGAGCTGCA 840
 913 TTCAGAGGACTTCAGAGGACGCTTGGGGAGCTTGGGGTCTCTGAGACCTGAGGAGCTGCA 972
 841 GGCCTTGCTGGGGTACCAAGGATGCGAGGCCCAAGAGGGCCCCCGGCTCTCTGGCCCA 900

|||||
 Db 973 GGCCTTGCTGGGGTACACAGCATGCGAGGCCCAAGAGGCCCGCCGCTCTGGGCCCA 1032
 901 TCAGAGCGGTGTGTGCTCCCTGCGCCCTGCAGATGAGCCACCCCGCAGGAGCAAT 960
 1033 TCAGAGCGGTGTGTGCTCCCTGCGCCCTGCAGATGAGCCACCCCGCAGGAGCAAT 1092
 961 GGCTGCGCCGCTCAGTGAAGAACTTCACAGCAAAATGCTACTATTCTTTCAGTTGAGAA 1020
 1093 AGCTGCGCCGCTCAGTGAAGAACTTCACAGCAAAATGCTACTATTCTTTCAGTTGAGAA 1152
 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGCAAGCTTTCACATCTTTTTCATA 1080
 1153 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGCAAGCTTTCACATCTTTTTCATA 1212
 1081 AACACTAGAGAGAACAGCAATGATATAAAAAACAGATGTGTAGGAGAGAGCCACTGG 1140
 1213 AACACTAGAGAGAACAGCAATGATATAAAAAACAGATGTGTAGGAGAGAGCCACTGG 1272
 1141 ATCGGCTCAGACACTCAGAGGCTGAATGAATGAATGAGTGGCTGGATGGACATCTCA 1200
 1273 ATCGGCTCAGACACTCAGAGGCTGAATGAATGAATGAGTGGCTGGATGGACATCTCA 1332
 1201 GACTACAAAATTTGAAAGCTGAGACAGCCGATTAATCTGGGCTCATGCGCATGGCCAGGA 1260
 1333 GACTACAAAATTTGAAAGCTGAGACAGCCGATTAATCTGGGCTCATGCGCATGGCCAGGA 1392
 1261 GAAGACTGTGTGGTGTATTTATGCTGGCGAGTGAACAGATTTCCATGTGAAGAGCTG 1320
 1393 GAAGACTGTGTGGTGTATTTATGCTGGCGAGTGAACAGATTTCCATGTGAAGAGCTG 1452
 1321 AATACTTATTTGGCAAAAGACAGAGAGACAGTACTGTCAATCTCATTTA 1371
 1453 AATACTTATTTGGCAAAAGACAGAGAGACAGTACTGTCAATCTCATTTA 1503

RESULT 5
 AX490920 2005 bp DNA linear PAT 16-AUG-2002
 LOCUS

DEFINITION Sequence 27 from Patent WO0200690.
 ACCESSION AX490920
 VERSION AX490920.1 GI:22323797
 KEYWORDS

SOURCE
 ORGANISM
 human.
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1
 Baker,K.P., Ferrara,N., Gerber,H., Gerlitsen,M.E., Goddard,A.,
 Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
 Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I.

and Ye,W.
 Compositions and methods for the diagnosis and treatment of
 disorders involving angiogenesis

JOURNAL
 Patent: WO 0200690-A 27 03-JAN-2002;
 Genentech, Inc. (US)

FEATURES
 source location/Qualifiers
 1..2005
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 606 a 493 c 491 g 415 t
 ORIGIN

Query Match 99.0%; Score 1357.4; DB 6; Length 2005;
 Best Local Similarity 99.9%; Pred. No. 3.2e-305;
 Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 ATGAACAGCCAGCTCACTCAATTCACAGAGTGCAGATGAGACATCAACCACTATCTCTCAA 60
 134 ATGAACAGCCAGCTCACTCAATTCACAGAGTGCAGATGAGACATCAACCACTATCTCTCAA 193
 61 GCCAAGAGGAGACCTGAAAGACCTGCAGAGCTTACACAAAGATGACAGATAGACAA 120

Db 194 GCCACGAGCAGAACTGGAAGACCTGCAGACTTACACAAAGATGCAGATAGACA 253
 QY 121 GCCATCAAGTTCAACCAACTGGAGGAAAGCTTCAGCTTTGAGAGCGATATTGTGAA 180
 Db 254 GCCATCAAGTTCAACCAACTGGAGGAAAGCTTCAGCTTTGAGAGCGATATTGTGAA 313
 QY 181 ATCATTAAGCAATATCATGTTTACACAGCCACCACCTGGGGAGCGTGCACCAATCTTAAT 240
 Db 314 ATCATTAAGCAATATCATGTTTACACAGCCACCACCTGGGGAGCGTGCACCAATCTTAAT 373
 QY 241 GAAGTCAGAGCAGCTTGCACAGATACCTTACCAAAACACAGATGATGTGACCTCTTG 300
 Db 374 GAAGTCAGAGCAGCTTGCACAGATACCTTACCAAAACACAGATGATGTGACCTCTTG 432
 QY 301 AATAATACCTGTGGCCAAACATCCGTTTGATTCGTTTCTCTCAGAGATGACAAGATTGG 360
 Db 433 AATAATACCTGTGGCCAAACATCCGTTTGATTCGTTTCTCTCAGAGATGACAAGATTGG 492
 QY 361 ATGAGGTCGAGGTTTACACACTGAAAGTACCACTTATCAGTGAATGGAAGAATGAAG 420
 Db 493 ATGAGGTCGAGGTTTACACACTGAAAGTACCACTTATCAGTGAATGGAAGAATGAAG 552
 QY 421 CTGTAGACTCCAAAGCAGTGTGCTCAGCTCATCAGAAATTTTCAATPACTACAAGGTCACG 480
 Db 553 CTGTAGACTCCAAAGCAGTGTGCTCAGCTCATCAGAAATTTTCAATPACTACAAGGTCACG 612
 QY 481 GGGCCGAGGGGTCCAAGAGGTGACAGAGATCCCAAGGAGCCCTGTGGCCCACTGTGGCAAC 540
 Db 613 GGGCCGAGGGGTCCAAGAGGTGACAGAGATCCCAAGGAGCCCTGTGGCCCACTGTGGCAAC 672
 QY 541 AAGGAGACGAAAGAGAGAAAGGGGAGCCTTGAGACCACTGTGGCCCTGTGGGTTGAGAGAGC 600
 Db 673 AAGGAGACGAAAGAGAGAAAGGGGAGCCTTGAGACCACTGTGGCCCTGTGGGTTGAGAGAGC 732
 QY 601 CCAATTGACACAGCTGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCACG 660
 Db 733 CCAATTGACACAGCTGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCACG 792
 QY 661 GGGCCGAGGGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 Db 793 GGGCCGAGGGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
 QY 721 CCAAGGCCCCCGGGGCCCCCAGGAGGAAAGAGGACTCCCGGCCCCCTCAGGGCCCTCCTGCGC 780
 Db 853 CCAAGGCCCCCGGGGCCCCCAGGAGGAAAGAGGACTCCCGGCCCCCTCAGGGCCCTCCTGCGC 912
 QY 781 TTCCAGGAGACTTTCAGGAGCAGCTTGGGAGAGCTGCGGAGTGCCTGCGGAGACTGCGCA 840
 Db 913 TTCCAGGAGACTTTCAGGAGCAGCTTGGGAGAGCTGCGGAGTGCCTGCGGAGACTGCGCA 972
 QY 841 GGGCTTGCTGGGGTACAGGAGATGCCAGGCCCAAGGGCCCCCGGCCCCCTCCTGCGCCA 900
 Db 973 GGGCTTGCTGGGGTACAGGAGATGCCAGGCCCAAGGGCCCCCGGCCCCCTCCTGCGCCA 1032
 QY 901 TCAGAGAGCGGTGGTGGCCCTGCGGCTTCAGAAATGAGCCAAACCCCGGAGCGGAGGACAT 960
 Db 1033 TCAGAGAGCGGTGGTGGCCCTGCGGCTTCAGAAATGAGCCAAACCCCGGAGCGGAGGACAT 1092
 QY 961 GGGTCGCGGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
 Db 1093 AGCTGCGCGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1152
 QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTTTCATA 1080
 Db 1153 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTTTCATA 1212
 QY 1081 AACACTAGAGAGAGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 Db 1213 AACACTAGAGAGAGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1272
 QY 1141 ATGCGGCTCAGAGCTCAGAGCGGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 Db 1273 ATGCGGCTCAGAGCTCAGAGCGGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1332

QY 1201 GACTACAAAATTGGAAGCTGGACAGCCGATPACTGGGGTCATGGCCATGGCCAGGA 1260
 Db 1333 GACTACAAAATTGGAAGCTGGACAGCCGATPACTGGGGTCATGGCCATGGCCAGGA 1392
 QY 1261 GAAGACTGTGCTGGGTTGATTTATGCTGGGACATGGAGCAATTCGAATGTGAAGAGCT 1320
 Db 1393 GAAGACTGTGCTGGGTTGATTTATGCTGGGACATGGAGCAATTCGAATGTGAAGAGCT 1452
 QY 1321 AATACTTCATTTGCGAAAAAGACAGAGAGACAGTACTGTCATCTCATTA 1371
 Db 1453 AATACTTCATTTGCGAAAAAGACAGAGAGACAGTACTGTCATCTCATTA 1503

RESULT 6
 HSM802985
 LOCUS
 DEFINITION Homo sapiens mRNA; cDNA DKFp547G1215 (from clone DKFp547G1215).
 ACCESSION AL713657
 VERSION AL713657.1 GI:19584339
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 Koehler, K., Beyer, A., Mewes, H.W., Weil, B. and Wiemann, S.
 1 (bases 1 to 1886)
 TITLE
 Direct Submission
 JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopperspitz 18a, D-82152

COMMENT
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
 This clone (DKFp547G1215) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de. Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cdna/>.
 Location/Qualifiers

FEATURES
 source
 1..1886

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="18p11.3"
 /clone="DKFp547G1215"
 /tissue_type="brain"
 /clone_lib="547 (synonym: hfbr1). Vector psport1; host DH10B; sites NotI + SalI"
 /dev_stage="fetal"
 1..1886
 /gene="DKFp547G1215"
 <1..1256
 /gene="DKFp547G1215"
 /note="similarity to asialoglycoprotein receptor (Mus musculus)"

CDS
 /product="hypothetical protein"
 /protein_id="CAD28466.1"
 /db_xref="GI:19584340"
 /translation="AIFKQLEPEROLFETDVLVNIISNISTAHHLRLTSLNLEVRT
 TCTDVLTKHTDLSLNNLTANI RUDSVSLRMQODIMSRDLTEVAANLSVIMEKLV
 DSKHQLIKNFILLOGPPGRPGRDGSGPPGTGNKGQGEKGEPPGPAGBERG
 PIGAPGPEGRGKSGKSGQKSGRSGSGPGSSSDPPGPPEKPEKGLPGQGP
 PFGLOGLEGVGPGRGLGPGVPMPPKPPGPPGSGAVPLALQNEPTPA
 PRDNCPPMKNFPTKCYFSEYKELFEDAKIFCDKSSSHLYFIFTRQQQIKKQW
 GRESHWIGLIDSERENKMLDGTSPDIKNMKRAGQPDMMGHGHEGDEAGLITAGQW
 NFGQEDVNNFICEKDRFTVSSAL"
 1838..1843
 /gene="DKFp547G1215"
 1869
 /gene="DKFp547G1215"

polyA_signal
 polyA_site
 BASE COUNT 569 a 444 c 445 g 428 t

Query Match		79.2%	Score 1086.2	DB 10	Length 2637
Best Local Similarity		87.0%	Pred. No. 4.2e-242		
Matches 1193		Conservative 0	Mismatches 178	Indels 0	Gaps 0
Qy	1	ATGAACAGCCAGCTCAACTCATTTCACTCAGAGTCCAGATGAGAGAACATCACTACTTCTCAA	60		
Db	947	ATGAATACCGACGCTCAGCTCATTTCAAGGTGAGTGAACAACTTAACCACTATCTCACAG	1006		
Qy	61	GCCAAAGCAGAACCTTGAAGACCTGCCAGACTTACACAAAGATGACAGATAGACA	120		
Db	1007	GCCAAAGCAGAGCCTGAAGACCTTGACGACTTACACAAAGATACAAATAATAGAAACA	1066		
Qy	121	GCCATCAGTTCAACCACTGSGAAGGCTTCACGCTCTTGAGAGGATATTTGTGAC	180		
Db	1067	GCTGCAAGTTCAAGCACTTGAGAACGCTTCCAGGTTCTTGAGCAATATTTGTGAC	1126		
Qy	181	ATCATTTACCAATATATAGTTACACAGCCCAACCACTGCGGAGCGCTGACCGCATCTAAAT	240		
Db	1127	ATCATTTACCAACATATAGCTACACACCCCACTACCTGAGSACACTGACCGCATCTGAAT	1186		
Qy	241	GAAGTCAGGACCACTTGCAAGATACCTTACCAAAACACACAGATGATCTGACCTCTTG	300		
Db	1187	GATGTTAGGACCACTATGACACAGACACTTTGACCAACACACGAGATGACTGACCTCTTG	1246		
Qy	301	AATATATACCTTGCCCAACATCCGTTTGATTTCTGTGTTTCTCTAGGATGCAACAAGTTTG	360		
Db	1247	AATATACACTATAGTCAACATCCGTTTGATTTCTGTGTTTCTCTAGGATGCAAGAACATG	1306		
Qy	361	ATGAGTGTGAGGTTATGACACTGAACTAGCACTTATCACTAGTATATGAAACAAATGAG	420		
Db	1367	ATGAGTGTCAAGTTATGACACTGAACTAGTGGCACTTATCACTAGTATATGAAAGATGAAA	1366		
Qy	421	CTAGTATACCTCAAGCATGTGTAGCTCATCAAGAAATTTTACAATCACTCAAGGTCACCG	480		
Db	1367	CTGTTGATCTCCAAACAGGCTGAGCTCATCAAGAACTTTACCATTTCTCAAGAGTCCCTCT	1426		
Qy	481	GGCCCCAGGGGTCGAAGAGTGAACAGAGGATCCCAAGGAGCCCCCTGGCCCACTGGCCAC	540		
Db	1427	GGCCCCAGAGGTCGAAGAGTGAACAGAGGATCTTCAAGGAGCAACTGTGTCAACTGGCAAC	1486		
Qy	541	AAGGACAGAAAGAGAGAAAGGGGAGGCTGTGACCACTGCGCCCTGCGGGTGAAGAGGCG	600		
Db	1487	AAGGACAGAAAGAGAGAGAGAGGAGAGGAGCGCTGTCCACTGCGGCGCTGAGAGGGCG	1546		
Qy	601	CCAATTTGACCACTGTGTCCCCCGGAGAGCGTGGCGCAAAAGATCTAAAGGCTCCAC	660		
Db	1547	ACAATTTGACCACTGTGTCCCCCGGAGAGCGCTGTGAGAGCGCTGACCAAAAGATCTCAAG	1606		
Qy	661	GGCCCCAAAGGCTCCCGGTGTTCCCTGGGGAAGCCGCGCCCTCAGGGCCCCAGTGGGGAG	720		
Db	1607	GGTCCCAAGGATCTCGGTGGGTGCCCAAGGAAAGCTGGGCCCTCAAGGACCTTATGTGGGGAG	1666		
Qy	721	CCAGGCCCCCGGGGCCCAACAGGCAAAAGAGGAGACTCCCGGCGCTCAGAGGCCCTCTGTGC	780		
Db	1667	CCAGGACCAACCAAGTGTCAACAGGCAAAAGATGGAATCTCTGCGGCCCTCAGGGCCCTCTGTGC	1726		
Qy	781	TTTCAAGGCACTTCAAGGACCGCTGTGGGAGCCGTGGGGTGTGAGACTTCAGGGACTGCCA	840		
Db	1727	TTTCAAGGCACTTCAAGGACCACTGTGGGTGTGAGCTGTGAGACTTCGAGACTTCGGGGTTGCCA	1786		
Qy	841	GGCTTGCGCTGGGGGTACAGAGCAATGCCAGGGCCCCCAAGGGCCCCCGGCGCTCTGGGCCA	900		
Db	1787	GGCTTGCGCAAGGGGTGCCAGGCAATGCCCTGGGCTTAAAGGGCAACCTGTGGCCCTCAGAGCCCC	1846		
Qy	901	TCAGAGAGGATGTGTGCTCCCTGTGCAGATGTAGCAACCCCGGACACGGAGAGCAAT	960		
Db	1847	TCAGAGAGCAATGTAGGACCAATGTGCTGTGCAGATGTAAACCAACCCCAAGATCAAGGTCAAC	1906		
Qy	961	GGCTGCGCGCTCACTGTGAAGAACTTCAACAGAAATGTCTACTATTTTTCAGTTGAGAA	1020		
Db	1907	GGATGTGCGCGCTCACTGTGAAGAACTTCAACAGATTAATGTCTACTATTTTTCAGTTGAGAAA	1966		
Qy	1021	GAATTTTTTGAGAGATGCAAGACTTTTCTGTGAGAGCAAGCTTTCACATCTGTGTTTCATA	1080		

[illegible]

BASE COUNT	1013 a	757 c	759 g	762 t
polyA_signal	3246.	3251		
ORIGIN	/gene="src1"			

Query Match	79.0%;	Score 1083;	DB 10;	Length 3291;
Best Local Similarity	86.9%;	Pred. No. 2.4e-241;		
Matches 1191; Conservative	0;	Mismatches 180;	Indels 0;	Gaps 0;

QY	1	ATTAAGCCGAGCTCAACTCATTTCAAGGTCAGATGAGAGACATATCCACTATCTCTCA	60
Db	932	ATGATATGCCAGCTCAGCTCTATTCAAGGTGAGATGGACAATTATCCATCTATCCACAG	991
QY	61	GCCACACGACGAGAACTGTAAAGACCTGCAGAGCTTATACAAAGATGACAGAAATGAGACA	120
Db	992	GCCAAAGCAGCAGACCTGTAAAGACCTTTCAGAGCTTATACAAAGATATACAGAAATATGAGACA	1051
QY	121	GCCATCAAGTTCAACCACTGAGGAGAGCGCTTCCAGCTCTTTGAGAGGGAATATGTGAA	180
Db	1052	GCTGTCAAGTTCCAGCAACTTGTAGGAACCTTCCAGGCTCTTTGAGACAGATATTTGTGAA	1111
QY	181	ATCATTTAGCAATTTACGTTATACAGGCCACACACTTCGAGAGCTTACCAAGATCTMAAT	240
Db	1112	ATCATTTAGCAACATTTACGTTATACAGGCCACACACTTCGAGAGCTTACCAAGATCTTGAAT	1171
QY	241	GAATCGACGACCACTTTCAGAGATACCTTTACCAACACACAGATATCTGACCTCGTG	300
Db	1172	GATTTTGGACACACATATGCACAGACCTTGTACCAAGACAGAGATGACTTGACTCTCTTG	1231
QY	301	AATATATCCCTGGCCAAATCCGTTTGGATTCTGTTTCTTCAGAGATCAACAGATTTG	360
Db	1232	AATATACACTATGTCAAACATCCGCTTGATTTCTATTTCTTCAGAGATGACAGACATG	1291
QY	361	ATGAGGTGAGGTTTAGACACTGAAATGACCAATTATCATGATATTTATGAAAGAAATGAG	420
Db	1292	ATGAGGTCAAAAGTTTAGACACTGTAAGTGGCCAACTTATCATGTTGTTATGGAAGAGATGAA	1351
QY	421	CTATATACATCCCAAGCATGGTTCAGCTCATCAAGAAATTTACAAATCTCAAGATCCACCG	480
Db	1352	CTGTTTACTCTCAAGACGCTGACTCATCAAGAAATTTACAAATCTCAAGATCTCTCT	1411
QY	481	GGCCCCAGGGGTCACAGAGGTGACAGAGATCCACAGGACCCCTCGGGCCCAACTGACAC	540
Db	1412	GGCCCCAGAGGTCACAAAGGTGACAGAGATCTCAAGGACACACCTGGTTCCAACTGGCAC	1471
QY	541	AAGGACAGAAAGAGAGAAAGGGGAGCCTGACACACTGGCCCTGGCGGTGAGAGAGC	600
Db	1472	AAGGACAGAAAGAGAGAAAGGGAGAGCCTGGTTCACCTGGCCCTGGCGGTGAGAGAGGC	1531
QY	601	CCATTTGACACAGCTGGTCCCGCCGGAGAGCGGGCGGCAAGATCTMAAGGCTCCAG	660
Db	1532	ACAATTTGACACAGCTGGCCCTCTGTGAGAGCGGTGGAGCAAAAGATCCAAAGGCTCACAG	1591
QY	661	GGCCCCAAAGGCTCCGTTGGTTCCCTCTGGAGAAACCGCGGCTTCAGAGGCCACAGTGGAGC	720
Db	1592	GGTCCCAAGATCTCGTGGTCCCGCAGGAAACCTGGCGCTTCAGAGACCTATGTTGGGAGC	1651
QY	721	CCAGGCCCCCGGGCCACACGACGCAAAAGAGGACTCCCGGCGCTTCAGAGGCCCTCTGGC	780
Db	1652	CCAGAGCACACCAAGTCCACACGACGCAAGATGAGATCTCCGCGCTTCAGAGGCCCTCTGGC	1711
QY	781	TTTCAGGAGCATTCAGGGCACCTGTGGGTGAGCCCTGGAGTACCTGGAGCTTCGGGGGTGCA	840
Db	1712	TTTCAGGAGCATTCAGGGCACCTGTGGGTGAGCCCTGGAGTACCTGGAGCTTCGGGGGTGCA	1771
QY	841	GGCTTTCCTGGGTACAGGCAATGCCAGGACCCCAAGAGGCCCCCGGGGCTCTTGAGCCA	900
Db	1772	GGCTTTCGCAAGGGGTGCCAGGCAATGCTCTGGGGCTTAAGAGACACACTGGCCCTTCAGGCC	1831
QY	901	TCAGAGAGGCTGTGCCCTGGCCCTTGCAAAATGAGCCACCCCGGACCGAGAGACAT	960

Db	1832	TCAGAGAGCATGGAGGCCATTGGCTCTGCGAGATGAAACCAACCCACAGCATGAGGTCAAC	1891
Qy	961	GGCTGCCCGCCTCACTCGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAA	1020
Db	1892	GGATGTCCGCCTCACTCGAAGAACTTCACAGATAAATGCTACTATTTTTCATTGGAAAA	1951
Qy	1021	GAATTTTGTGAGATGCAAAAGCTTTCTGTGAAGACAGCTTTCACATCTTGTTCATA	1080
Db	1952	GAATTTTGTGAGATGCTAAGCTTTTCTGTGAAGACAAATCTCCATCTCTTTCATA	2011
Qy	1081	AACCTGAGAGGAGCAACGCAATGGATPAAAAAACAATGTTGGGAGAGAGGCCACAG	1140
Db	2012	AACCTAAGAGAGAACGCAATGGATPAAAAACATATCCGTGGGGAGAGAAAGCCATTGG	2071
Qy	1141	ATCGGCGCTCAGACACTAGAGCGCTGAAAAATGAATGAAATGGCTGGATGGACATCTCCA	1200
Db	2072	ATCGGCGCTCAGACACTAGAAACAGAAAGAGATGAAATGGCTAGACGGGTCACTCTT	2131
Qy	1201	GACTPACAAAAATTTGAAAGCTGACAGCCGGATTACTGGGGTCTATGGCCATGGCCACGA	1260
Db	2132	GATTPACAAAAATCTGAAAGCTGGACAAACAGATAACTGGGGCAGTAGGCCATGGCCACGA	2191
Qy	1261	GAAAGCTGTGCTGGGTTGATTATGCTGGGCACTGGAACGATTTCCATATGTAAGACTC	1320
Db	2192	GAAAGCTGTGCTGGCTTGATTATGCGAGAGCACTGGAAAGACTTCCAGTGTGATTAATTC	2251
Qy	1321	AATACCTTCATTGTCGAAAAAGAGGAGAGAGCTACTGTCCTCATCTGCATTTA	1371
Db	2252	AATACCTTCATTGTCGAAAGGAAAGGAGAGCGCATGTATCATCTATTCATTTA	2302

```

RESULT 9
LOCUS      AB052103
DEFINITION Homo sapiens SRCU mRNA for scavenger receptor with C-type lectin
            type II, complete cds.
ACCESSION  AB052103
VERSION    AB052103.1
KEYWORDS   GI:13365552
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (sites)
AUTHORS    Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE      Molecular cloning and functional characterization of a human
            scavenger receptor with C-type lectin (SRCU), a novel member of a
            scavenger receptor family
JOURNAL    Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE    21092718
REFERENCE  2 (bases 1 to 4330)
AUTHORS    Nakamura,K. and Nakamura,T.
TITLE      Direct Submission
JOURNAL    Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
            School of Medicine, Division of Biochemistry, Biomedical Research
            Center: 2-2, Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:nakamuraonbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
            Fax:81-6-6879-3789)
FEATURES   Location/Qualifiers
            1..4330
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               1..4330
               /gene="SRCU"
               3..1871
               /gene="SRCU"
               /codon_start=1
               /product="scavenger receptor with C-type lectin type II"
               /protein_id="BAB39148.1"
               /db_xref="GI:13365553"
               /translation="MKDDFAEEBEPSPFGKFRBEIHGTQCPKCHNNWALKFSITLLY
               ICLNLTITVAIVIGYKVAEKMDANTGVGKERSRQRYDYDKLVAVESDLKLDIDQGTCKAI

```

STNSELSTFRSDILDRQOLREITEXTSKNKDLEKLSAGALVDROSQLEETLENN
 SFLITTVNKTLDAYNGYVYNTLQODTSVLGNLQNMYSNHYVIMNINLNTLOVOORN
 LITNLSRVNDSOAIYORINDFOINQOYDMLKEVOJLQJOTLAORLAAKNOLAE
 KANNDELDMNSQINSEFGOMENITITISANBONELKDLDIRHDANPRAIKENOLEE
 RQIPELTVNITISNTSYTAHLRTLSNTEPCTPLTRKPDLDLSINLTANI
 RDSVSLRMOODLMRSRLDTEVANLSVIEBMLVDSKQOLIKNTTILQPPGPGP
 RDRSGSQGPPGPGPGGQKKEKEPEPAGRGIGIAGPAGPEGGSGSKSQGPK
 GSGSGPKGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG
 LPQVPMPPKPGPPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG
 BASE COUNT 1186 a 1065 c 1035 g 1044 t
 ORIGIN

Query Match 70.1%; Score 960.4; DB 9; Length 4330;
 Best Local Similarity 99.9%; Pred. No. 8.2e-213;
 Matches 961; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

1 ATGACAGCAGCAGCTCATCTATTCAGAGTCAGATGAGAAACATCACCATTCTCCAA 60
 858 ATGACAGCAGCAGCTCATCTATTCAGAGTCAGATGAGAAACATCACCATTCTCCAA 917
 61 GCCAAGCAGCAGAACTGAAAGACCTCAGAGCTTACCAAAAGATGAGAGAAATAGACA 120
 918 GCCAAGCAGCAGAACTGAAAGACCTCAGAGCTTACCAAAAGATGAGAGAAATAGACA 977
 121 GCCATCAAGTTCAACCACTGAGAGAACCTTCCAGCTTTTGAGAGCATTTGTGAAC 180
 978 GCCATCAAGTTCAACCACTGAGAGAACCTTCCAGCTTTTGAGAGCATTTGTGAAC 1037
 181 ATCATGACATATCATGTTACAGAGCCACACCTGGGAGCGTACAGACATCTAAT 240
 1038 ATCATGACATATCATGTTACAGAGCCACACCTGGGAGCGTACAGACATCTAAT 1097
 241 GAACTCAGAGCACTTCAAGATACCTTACCAAAACACAGATGATGACCTCCTG 300
 1098 GAACTCAGAGCACTTCAAGATACCTTACCAAAACACAGATGATGACCTCCTG 1157
 301 AATATATCCTGCGCAGCATCCCTTTGGATTCCTCTCAGAGATCAGCAAGATTTG 360
 1158 AATATATCCTGCGCAGCATCCCTTTGGATTCCTCTCAGAGATCAGCAAGATTTG 1217
 361 ATGAGTCGAGAGTTAGACATGAGTGAAGTCACTTATGATGATGAGAAATGAAG 420
 1218 ATGAGTCGAGAGTTAGACATGAGTGAAGTCACTTATGATGATGAGAAATGAAG 1277
 421 CTAGTAGACTCCAGCATGTGCTCATCAAGATTTTACAATATCTACAAGTCCACCG 480
 1278 CTAGTAGACTCCAGCATGTGCTCATCAAGATTTTACAATATCTACAAGTCCACCG 1337
 481 GGGCCGAGGGGTCACAAGAGTGAAGAGATCCAGAGACCCCTGGCCCAACTGGCAAC 540
 1338 GGGCCGAGGGGTCACAAGAGTGAAGAGATCCAGAGACCCCTGGCCCAACTGGCAAC 1397
 541 AAGGACAGAAAGAGAGAGAGGGGAGCTGAGACCACTGGCCCTGGGGTGAAGAGGC 600
 1398 AAGGACAGAAAGAGAGAGAGGGGAGCTGAGACCACTGGCCCTGGGGTGAAGAGGC 1457
 601 CCAATTGACACAGCTGTGCTCCCGGAGAGCTGCGCAAAAGATCTAAAGGCTCCAG 660
 1458 CCAATTGACACAGCTGTGCTCCCGGAGAGCTGCGCAAAAGATCTAAAGGCTCCAG 1517
 661 GGGCCCAAAAGGCTGCTGTTCCCTGGGAGACCCGCGCTCAGGGCCCCCAGTGGGAG 720
 1518 GGGCCCAAAAGGCTGCTGTTCCCTGGGAGACCCGCGCTCAGGGCCCCCAGTGGGAG 1577
 721 CAGAGCCCCCGGAGCAGAGCAAGAGAGCACTCCCGGCTCAGGGCCCTCCTGGG 780
 1578 CAGAGCCCCCGGAGCAGAGCAAGAGAGCACTCCCGGCTCAGGGCCCTCCTGGG 1637
 781 TTCAGAGGACTTACAGGACCTGTTGGGAGCTTGGGGTCTGAGACCTCGGGGACTGCCA 840
 1638 TTCAGAGGACTTACAGGACCTGTTGGGAGCTTGGGGTCTGAGACCTCGGGGACTGCCA 1697
 841 GGCCTGCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCTCCTCGGCCCA 900

Db 1698 GGGCTTGGCTGGGTACAGGACATGACAGGCCCAAGGGCCCCCGGCTCCTCGGCCCA 1757
 QY 901 TCAGGAGCGGTGGTGGCCCCCGGCGCTGACAGATGAGCAACCCCGGACCGGAGCAAT 960
 Db 1758 TCAGGAGCGGTGGTGGCCCCCGGCGCTGACAGATGAGCAACCCCGGACCGGAGCAAT 1817
 QY 961 GG 962
 Db 1818 AG 1819

RESULT 10
 AC016128/c
 LOCUS
 DEFINITION Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
 SEQUENCE 8 unordered pieces.
 ACCESSION AC016128
 VERSION AC016128.4 GI:10046526
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 169088)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 18, clone RP11-324G2
 Unpublished
 2 (bases 1 to 169088)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donnell,L., Doyle,M.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Margulis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testfay,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 COMMENT
 Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 9, 2000 this sequence version replaced q1:6649269.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L999
 GenBank accession: 324_G-2
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye, 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 165522 bases at least Q40
 Consensus quality: 167248 bases at least Q30
 Consensus quality: 167996 bases at least Q20
 Insert size: 17000; agarose-fp
 Insert size: 168388; sum-of-ctrls
 Quality coverage: 6.2 in Q20 bases; agarose-fp
 Quality coverage: 6.5 in Q20 bases; sum-of-ctrls

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 17943: contig of 17943 bp in length
 * 17944 18043: gap of 100 bp
 * 18044 22424: contig of 4381 bp in length
 * 22425 22524: gap of 100 bp
 * 22525 38094: contig of 15570 bp in length
 * 38095 38194: gap of 100 bp
 * 38195 49220: contig of 11026 bp in length
 * 49221 49320: gap of 100 bp
 * 49321 66394: contig of 17074 bp in length
 * 66395 66494: gap of 100 bp
 * 66495 91692: contig of 25198 bp in length
 * 91693 91792: gap of 100 bp
 * 91793 120869: contig of 29077 bp in length
 * 120870 120970: gap of 100 bp
 * 120970 169088: contig of 48119 bp in length.

FEATURES
 source
 1..169088
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-324G2"
 /clone_lib="RP11 Human Male BAC"
 1..17943
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"

misc_feature
 18044..22424
 /note="assembly_fragment"
 22525..38094
 /note="assembly_fragment"
 38195..49220
 /note="assembly_fragment"
 49321..66394
 /note="assembly_fragment"
 66495..91692
 /note="assembly_fragment"
 91793..120869
 /note="assembly_fragment"
 120970..169088
 /note="assembly_fragment"
 clone_end:TY
 vector_side:right"

BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
 ORIGIN

Query Match 35.7%; Score 488.8; DB 2; Length 169088;
 Best Local Similarity 99.6%; Pred. No. 5.8e-103;
 Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 471 AGGTCACCGGGCCCGAGGCTGTCACAGAGTGCACAGAGTCCACAGGACCCCTGGGCC 530
 |||||||
 Db 153817 AGGTCACCGGGCCCGAGGCTGTCACAGAGTGCACAGAGTCCACAGGACCCCTGGGCC 153758
 |||||||
 QY 531 AACTGGCAACAAGGACAGAAAGAGAGAGAGGAGGAGCCTGACACACTTGCCCTGCGGG 590
 |||||||
 Db 153757 AACTGGCAACAAGGACAGAAAGAGAGAGAGGAGGAGCCTGACACACTTGCCCTGCGGG 153698
 |||||||
 QY 591 TGAAGAGAGCCCAATTGGACACAGCTGTGTCCTCCCGAGAGCGTGGCGGCAAAAGATCTAA 650
 |||||||
 Db 153697 TGAAGAGAGCCCAATTGGACACAGCTGTGTCCTCCCGAGAGCGTGGCGGCAAAAGATCTAA 153638
 |||||||
 QY 651 AGGCTCCAGGGCCCAAGAGCTCCCGTGTTCCTCGGGAAGCCCGGCTTCAGGGGCC 710
 |||||||
 Db 153637 AGGCTCCAGGGCCCAAGAGCTCCCGTGTTCCTCGGGAAGCCCGGCTTCAGGGGCC 153578
 |||||||
 QY 711 CAGTGGGAGCCCAAGGCCCCCGGGCCACACAGGCAAAAGAGGACTCCCGGCTTCAGGG 770
 |||||||

Db 153577 CAGTGGGAGCCCAAGGCCCCCGGGCCACACAGGCAAAAGAGGACTCCCGGCTTCAGGG 153518
 QY 771 CCCTCTGGCTTCACAGGACTTCAGGGACCGCTGGGGAGCCCGGGGTGCTTGAGACTCG 830
 |||||||
 Db 153517 CCCTCTGGCTTCACAGGACTTCAGGGACCGCTGGGGAGCCCGGGGTGCTTGAGACTCG 153458
 |||||||
 QY 831 GGGACTGCCAGGCTTCCTGGGGTACACAGGATGACAGCCGCCCAAGGGCCCCCGGGCC 890
 |||||||
 Db 153457 GGGACTGCCAGGCTTCCTGGGGTACACAGGATGACAGCCGCCCAAGGGCCCCCGGGCC 153398
 |||||||
 QY 891 TCCTGGCCATCAGAGCGGCTGTGCTCCCTGGCCCTGCAGATGAGCAACCCCGGCACC 950
 |||||||
 Db 153397 TCCTGGCCATCAGAGCGGCTGTGCTCCCTGGCCCTGCAGATGAGCAACCCCGGCACC 153338
 |||||||
 QY 951 GGAGGACAAATGG 962
 |||||||
 Db 153337 GGAGGACAAATGG 153326

RESULT 11
 AP000915/c 188439 bp DNA linear PRI 27-APR-2002
 LOCUS Homo sapiens genomic DNA, chromosome 18p, clone:RP11-720L2, complete
 DEFINITION sequences.
 ACCESSION AP000915
 VERSION AP000915.5 GI:20334314
 KEYWORDS HTG.
 SOURCE Homo sapiens DNA, clone:RP11-720L2.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Homo sapiens genomic DNA
 JOURNAL Published Only in Database (1999)
 REFERENCE 2 (bases 1 to 188439)
 AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: hattori@psc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 On Apr 26, 2002 this sequence version replaced gi:9188470.

COMMENT
 FEATURES
 Location/Qualifiers
 source
 1..188439
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="18p"
 /chromosome="18"
 /clone="RP11-720L2"

BASE COUNT 54003 a 41134 c 40027 g 53275 t
 ORIGIN

Query Match 35.7%; Score 488.8; DB 9; Length 188439;
 Best Local Similarity 99.6%; Pred. No. 5.8e-103;
 Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 471 AGGTCACCGGGCCCGAGGCTGTCACAGAGTGCACAGAGTCCACAGGACCCCTGGGCC 530
 |||||||
 Db 44938 AGGTCACCGGGCCCGAGGCTGTCACAGAGTGCACAGAGTCCACAGGACCCCTGGGCC 44879
 |||||||
 QY 531 AACTGGCAACAAGGACAGAAAGAGAGAGAGGAGGAGCCTGACACACTTGCCCTGCGGG 590
 |||||||
 Db 44878 AACTGGCAACAAGGACAGAAAGAGAGAGAGGAGGAGCCTGACACACTTGCCCTGCGGG 44819
 |||||||
 QY 591 TGAAGAGAGCCCAATTGGACACAGCTGTGTCCTCCCGAGAGCGTGGCGGCAAAAGATCTAA 650
 |||||||
 Db 44818 TGAAGAGAGCCCAATTGGACACAGCTGTGTCCTCCCGAGAGCGTGGCGGCAAAAGATCTAA 44759
 |||||||
 QY 651 AGGCTCCAGGGCCCAAGAGCTCCCGTGTTCCTCGGGAAGCCCGGCTTCAGGGGCC 710
 |||||||

```

Db 44758 AGCTCCAGAGCCGCCAAAGCTCCCTGCTCCCTCGAGGAGCCGCCCTCAGGGCTC 44699
OY 711 CAGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGACTCCCGGGCCCTCAGGG 770
Db 44698 CAGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGACTCCCGGGCCCTCAGGG 44639
OY 771 CCCTCTGCTTCAGGAGCTTCAGGGACCGTGGGGAGCTGGGGTGCCTGAGCTCG 830
Db 44638 CCCCTCTGCTTCAGGAGCTTCAGGGACCGTGGGGAGCTGGGGTGCCTGAGCTCG 44579
OY 831 GGGACTCCAGAGCTGCTGCTGGGGTACCAAGCATCCAGGCCCAAGGGCCCCCGGGCCC 890
Db 44578 GGGACTCCAGAGCTGCTGCTGGGGTACCAAGCATCCAGGCCCAAGGGCCCCCGGGCCC 44519
OY 891 TCCTGGCCCTTCAGAGCGGTGGTGGCCCTGAGCCCTCAGAAATGAGCCCAAGCCCGGAC 950
Db 44518 TCCTGGCCCTTCAGAGCGGTGGTGGCCCTGAGCCCTCAGAAATGAGCCCAAGCCCGAC 44459
OY 951 GGAGGACAAATGG 962
Db 44458 GGAGGACAAATGG 44447

```

```

RESULT 12
AC024368 71044 bp DNA linear HTG 28-FEB-2000
LOCUS Homo sapiens chromosome 11 clone RP11-179K3 map 11, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC024368.1 GI:7108157
VERSION HTG: HTGS_PHASE0.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 71044)
2 (bases 1 to 71044)
REFERENCE 1 (bases 1 to 71044)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
JOURNAL Chepel'Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
REFERENCE Dearlino,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
AUTHORS Feneator,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Klein,J., Lander,S., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,Y., Morrow,J., Naylor,J.,
Norton,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```

----- Project Information
Center project name: L6145
Center clone name: 179_K3

```

```

* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

1 698: contig of 698 bp in length
699 798: gap of 100 bp
799 1595: contig of 697 bp in length
1496 1595: gap of 100 bp
1596 2284: contig of 689 bp in length
2285 2384: gap of 100 bp
2385 3056: contig of 672 bp in length
3057 3156: gap of 100 bp
3157 3851: contig of 695 bp in length
3852 3951: gap of 100 bp
3952 4627: contig of 676 bp in length
4628 4727: gap of 100 bp
4728 5409: contig of 682 bp in length
5410 5509: gap of 100 bp
5510 6183: contig of 674 bp in length
6184 6283: gap of 100 bp
6284 6984: contig of 701 bp in length
6985 7084: gap of 100 bp
7085 7768: contig of 684 bp in length
7769 7868: gap of 100 bp
7869 8550: contig of 682 bp in length
8551 8650: gap of 100 bp
8651 9352: contig of 702 bp in length
9353 9452: gap of 100 bp
9453 10140: contig of 688 bp in length
10141 10240: gap of 100 bp
10241 10922: contig of 682 bp in length
10923 11022: gap of 100 bp
11023 11709: contig of 687 bp in length
11710 11809: gap of 100 bp
11810 12511: contig of 702 bp in length
12512 12611: gap of 100 bp
12612 13304: contig of 693 bp in length
13305 13404: gap of 100 bp
13405 14095: contig of 691 bp in length
14096 14195: gap of 100 bp
14196 14881: contig of 686 bp in length
14882 14981: gap of 100 bp
14982 15674: contig of 693 bp in length
15675 15774: gap of 100 bp
15775 16460: contig of 686 bp in length
16461 16560: gap of 100 bp
16561 17250: contig of 690 bp in length
17251 17350: gap of 100 bp
17351 18040: contig of 690 bp in length
18041 18140: gap of 100 bp
18141 18836: contig of 696 bp in length
18837 18936: gap of 100 bp
18937 19621: contig of 685 bp in length
19622 19721: gap of 100 bp
19722 20425: contig of 704 bp in length
20426 20525: gap of 100 bp
20526 21224: contig of 699 bp in length
21225 21324: gap of 100 bp
21325 22017: contig of 693 bp in length
22018 22117: gap of 100 bp
22118 22815: contig of 698 bp in length
22816 22915: gap of 100 bp

```

22816	23609:	contlg of 694 bp	in length
23610	23709:	gap of 100 bp	
23710	24402:	contlg of 693 bp	in length
24403	24502:	gap of 100 bp	
24403	25101:	contlg of 689 bp	in length
25192	25291:	gap of 100 bp	
25292	25961:	contlg of 690 bp	in length
25882	26081:	gap of 100 bp	
26082	26774:	contlg of 693 bp	in length
26775	27563:	contlg of 689 bp	in length
27664	27663:	gap of 100 bp	
27664	28355:	contlg of 692 bp	in length
28356	28455:	gap of 100 bp	
28456	29145:	contlg of 690 bp	in length
29146	29245:	gap of 100 bp	
29246	29943:	contlg of 698 bp	in length
29944	30043:	gap of 100 bp	
30044	30745:	contlg of 702 bp	in length
30746	30845:	gap of 100 bp	
30846	31535:	contlg of 690 bp	in length
31536	31635:	gap of 100 bp	
31636	32339:	contlg of 694 bp	in length
32330	32429:	gap of 100 bp	
32430	33114:	contlg of 685 bp	in length
33115	33214:	gap of 100 bp	
33215	33898:	contlg of 684 bp	in length
33899	33998:	gap of 100 bp	
33999	34694:	contlg of 696 bp	in length
34695	34794:	gap of 100 bp	
34795	35486:	contlg of 692 bp	in length
35487	35586:	gap of 100 bp	
35587	36292:	contlg of 706 bp	in length
36293	36392:	gap of 100 bp	
36393	37095:	contlg of 703 bp	in length
37096	37195:	gap of 100 bp	
37196	37872:	contlg of 677 bp	in length
37873	37972:	gap of 100 bp	
37973	38663:	contlg of 691 bp	in length
38664	38763:	gap of 100 bp	
38764	39449:	contlg of 686 bp	in length
39450	39549:	gap of 100 bp	
39550	40236:	contlg of 687 bp	in length
40237	40336:	gap of 100 bp	
40337	41025:	contlg of 689 bp	in length
41026	41125:	gap of 100 bp	
41126	41818:	contlg of 693 bp	in length
41819	41918:	gap of 100 bp	
41919	42608:	contlg of 690 bp	in length
42609	42708:	gap of 100 bp	
42709	43399:	contlg of 691 bp	in length
43400	43499:	gap of 100 bp	
43500	44169:	contlg of 670 bp	in length
44170	44269:	gap of 100 bp	
44270	44966:	contlg of 697 bp	in length
44967	45066:	gap of 100 bp	
45067	45750:	contlg of 684 bp	in length
45751	45850:	gap of 100 bp	
45851	46545:	contlg of 695 bp	in length
46546	46645:	gap of 100 bp	
46646	47336:	contlg of 691 bp	in length
47337	47436:	gap of 100 bp	
47437	48133:	contlg of 687 bp	in length
48124	48223:	gap of 100 bp	
48224	48899:	contlg of 676 bp	in length
48900	48999:	gap of 100 bp	
49000	49672:	contlg of 673 bp	in length
49673	49772:	gap of 100 bp	
49773	50465:	contlg of 693 bp	in length
50466	50565:	gap of 100 bp	
50566	51234:	contlg of 689 bp	in length
51235	51354:	gap of 100 bp	
51355	52062:	contlg of 708 bp	in length

Query Match	35.3%;	Score 484.6;	DB 2;	Length 71044;
Best Local Similarity	98.2%;	Pred. No. 5.2e-102;		
Matches 490;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;

```

* 52063 52162: gap of 100 bp
* 52163 52847: contig of 685 bp in length
* 52848 52947: gap of 100 bp
* 52848 53646: contig of 699 bp in length
* 53647 53746: gap of 100 bp
* 53747 54409: contig of 663 bp in length
* 54410 54509: gap of 100 bp

```

464 TACTACAAGGTCACCGGGGCCAGGGGTCCAAGAGGTGACACAGAGGATCCAGGGACCCC 523

Db 43599 TAGCAAAAGTCCACCGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCAGGACCCC 43658

524 CTGCCCACTGGCAACAAGGACAGAAAGGAGAGGAAGGGGAGCCTGGACCACCTGGCC 583

Db 43659 CTGCCCACTGGCAACAAGGACAGAAAGGAGAGAGGAGGAGCTGGACC 43718

584 CTGGGGTGAGAGAGGCCCAATTGGACCAGCTGTCCTCCCCCGAGAGCGTGCGGCAAG 643

Db 43719 CTGGGGTGAGAGAGGCCAGGGGAGCCAGCTGTCCTCCCGGAGAGCGTGCGGCAAG 43778

644 GATCTAAGGCTCCACAGGGCCCCAAGGCTCCCGTGGTTCCTGGGAAGCCCGGCCCTC 703

Db 43779 GATCTAAGGCTCCAGGGCCCAAGGCTCCCGTGTTCCCTGGGAAGCCGCGCCTC 43838

704 AGGCCCCAGTGGGGACCCAGGCCCCCGGGCCACCAGGCAAGGACTCCCGGCC 763

db 43839 AGGCTTCAGTGGGACCCAGGCCCCCGGGCCAAAGAGGGACTCCCGGGCC 43898

754 CTACAGGGCCCTCTGCTTTCAGGGACTTCAAGGCAACGTTGGGGAGCCCTGGGGTGCCTG 823

Dh 43899 CTACCCCTCCTTCCACCCTTCCTCACCTTTCCCCCCC 43058

|||||||

[illegible]

42050 42051 42052 42053 42054 42055 42056 42057 42058 42059 42060 42061 42062 42063 42064 42065 42066 42067 42068 42069 42070 42071 42072 42073 42074 42075 42076 42077 42078 42079 42080 42081 42082 42083 42084 42085 42086 42087 42088 42089 42090 42091 42092 42093 42094 42095 42096 42097 42098 42099 42100 42101 42102 42103 42104 42105 42106 42107 42108 42109 42110 42111 42112 42113 42114 42115 42116 42117 42118 42119 42120 42121 42122 42123 42124 42125 42126 42127 42128 42129 42130 42131 42132 42133 42134 42135 42136 42137 42138 42139 42140 42141 42142 42143 42144 42145 42146 42147 42148 42149 42150 42151 42152 42153 42154 42155 42156 42157 42158 42159 42160 42161 42162 42163 42164 42165 42166 42167 42168 42169 42170 42171 42172 42173 42174 42175 42176 42177 42178 42179 42180 42181 42182 42183 42184 42185 42186 42187 42188 42189 42190 42191 42192 42193 42194 42195 42196 42197 42198 42199 42200 42201 42202 42203 42204 42205 42206 42207 42208 42209 42210 42211 42212 42213 42214 42215 42216 42217 42218 42219 42220 42221 42222 42223 42224 42225 42226 42227 42228 42229 42230 42231 42232 42233 42234 42235 42236 42237 42238 42239 42240 42241 42242 42243 42244 42245 42246 42247 42248 42249 42250 42251 42252 42253 42254 42255 42256 42257 42258 42259 42260 42261 42262 42263 42264 42265 42266 42267 42268 42269 42270 42271 42272 42273 42274 42275 42276 42277 42278 42279 42280 42281 42282 42283 42284 42285 42286 42287 42288 42289 42290 42291 42292 42293 42294 42295 42296 42297 42298 42299 42300 42301 42302 42303 42304 42305 42306 42307 42308 42309 42310 42311 42312 42313 42314 42315 42316 42317 42318 42319 42320 42321 42322 42323 42324 42325 42326 42327 42328 42329 42330 42331 42332 42333 42334 42335 42336 42337 42338 42339 42340 42341 42342 42343 42344 42345 42346 42347 42348 42349 42350 42351 42352 42353 42354 42355 42356 42357 42358 42359 42360 42361 42362 42363 42364 42365 42366 42367 42368 42369 42370 42371 42372 42373 42374 42375 42376 42377 42378 42379 42380 42381 42382 42383 42384 42385 42386 42387 42388 42389 42390 42391 42392 42393 42394 42395 42396 42397 42398 42399 42400 42401 42402 42403 42404 42405 42406 42407 42408 42409 42410 42411 42412 42413 42414 42415 42416 42417 42418 42419 42420 42421 42422 42423 42424 42425 42426 42427 42428 42429 42430 42431 42432 42433 42434 42435 42436 42437 42438 42439 42440 42441 42442 42443 42444 42445 42446 42447 42448 42449 42450 42451 42452 42453 42454 42455 42456 42457 42458 42459 42460 42461 42462 42463 42464 42465 42466 42467 42468 42469 42470 42471 42472 42473 42474 42475 42476 42477 42478 42479 42480 42481 42482 42483 42484 42485 42486 42487 42488 42489 42490 42491 42492 42493 42494 42495 42496 42497 42498 42499 42500 42501 42502 42503 42504 42505 42506 42507 42508 42509 42510 42511 42512 42513 42514 42515 42516 42517 42518 42519 42520 42521 42522 42523 42524 42525 42526 42527 42528 42529 42530 42531 42532 42533 42534 42535 42536 42537 42538 42539 42540 42541 42542 42543 42544 42545 42546 42547 42548 42549 42550 42551 42552 42553 42554 42555 42556 42557 42558 42559 42560 42561 42562 42563 42564 42565 42566 42567 42568 42569 42570 42571 42572 42573 42574 42575 42576 42577 42578 42579 42580 42581 42582 42583 42584 42585 42586 42587 42588 42589 42590 42591 42592 42593 42594 42595 42596 42597 42598 42599 42600 42601 42602 42603 42604 42605 42606 42607 42608 42609 42610 42611 42612 42613 42614 42615 42616 42617 42618 42619 42620 42621 42622 42623 42624 42625 42626 42627 42628 42629 42630 42631 42632 42633 42634 42635 42636 42637 42638 42639 42640 42641 42642 42643 42644 42645 42646 42647 42648 42649 42650 42651 42652 42653 42654 42655 42656 42657 42658 42659 42660 42661 42662 42663 42664 42665 42666 42667 42668 42669 42670 42671 42672 42673 42674 42675 42676 42677 42678 42679 42680 42681 42682 42683 42684 42685 42686 42687 42688 42689 42690 42691 42692 42693 42694 42695 42696 42697 42698 42699 42700 42701 42702 42703 42704 42705 42706 42707 42708 42709 42710 42711 42712 42713 42714 42715 42716 42717 42718 42719 42720 42721 42722 42723 42724 42725 42726 42727 42728 42729 42730 42731 42

[illegible][illegible][illegible][illegible]

DU 44V/3 CAGCACCGAGGACAAAGG 44V3/

RESULT 13

AP000900	LOCUS	AP000900	178022 bp	DNA	linear	HTG 30-MAY-2

DEFINITION Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKING DRAFT SEQUENCE, 20 unordered pieces.

ACCESSION	AP000900
VERSION	AP000900.3
	GI:8119043

KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	HOMO sapiens DNA clone: RP11-6683.111

ORGANISM	Homo sapiens
Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:	

REFERENCE
1 (bases 1 to 178022)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiwara, A., Yada, T., Totsuki, Y., Watanabe, H. and Sakaki, Y.

TITLE Homo sapiens 178,022 genomic DNA of 18p11.3
TOURNAMENT Published Online 12 Dec 2008 (1000)

REFERENCE 2 (bases 1 to 178022)

Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physi

and Chemical Research (KIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

COMMENT

Japan (E-mail:hatorigsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced g1:6997403.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hatorigsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center Clone name: RP11-683J11
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161944 bases at least Q40
Consensus quality: 170357 bases at least Q30
Consensus quality: 174322 bases at least Q20
Insert size: 176122; sum-of-contrigs
Quality coverage: 4.56x in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently consists of
20 contrigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contrigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

```
1 24283 contrig of 24283 bp in length
24384 46875 contrig of 22492 bp in length
46976 65058 contrig of 18083 bp in length
65159 81322 contrig of 16164 bp in length
81423 98823 contrig of 17401 bp in length
98924 110068 contrig of 11145 bp in length
110169 122716 contrig of 12548 bp in length
122817 133089 contrig of 10273 bp in length
133190 141206 contrig of 8017 bp in length
141307 149050 contrig of 7744 bp in length
149151 156206 contrig of 6955 bp in length
156206 162048 contrig of 5843 bp in length
162149 167480 contrig of 2262 bp in length
167581 169986 contrig of 2406 bp in length
170087 171909 contrig of 1823 bp in length
172010 172317 contrig of 308 bp in length
172418 174663 contrig of 2246 bp in length
174764 176271 contrig of 1508 bp in length
176372 178022 contrig of 1651 bp in length

Sequence updated (24-Dec-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 24283: contrig of 24283 bp in length
24284 24383: gap of 100 bp
24384 46875: contrig of 22492 bp in length
46876 46975: gap of 100 bp
46976 65058: contrig of 18083 bp in length
65059 65158: gap of 100 bp
65159 81322: contrig of 16164 bp in length
81323 81422: gap of 100 bp
81423 98823: contrig of 17401 bp in length
98824 98923: gap of 100 bp
98924 110068: contrig of 11145 bp in length
110069 110168: gap of 100 bp
110169 122716: contrig of 12548 bp in length
122717 122816: gap of 100 bp
```

```
* 122817 133089: contrig of 10273 bp in length
* 133090 133189: gap of 100 bp
* 133190 141206: contrig of 8017 bp in length
* 141207 141306: gap of 100 bp
* 141307 149050: contrig of 7744 bp in length
* 149051 149150: gap of 100 bp
* 149151 156105: contrig of 6955 bp in length
* 156106 156205: gap of 100 bp
* 156206 162048: contrig of 5843 bp in length
* 162049 162148: gap of 100 bp
* 162149 165118: contrig of 2970 bp in length
* 165119 165218: gap of 100 bp
* 165219 167480: contrig of 2262 bp in length
* 167481 167580: gap of 100 bp
* 167581 169986: contrig of 2406 bp in length
* 169987 170086: gap of 100 bp
* 170087 171909: contrig of 1823 bp in length
* 171910 172009: gap of 100 bp
* 172010 172317: contrig of 308 bp in length
* 172318 172417: gap of 100 bp
* 172418 174663: contrig of 2246 bp in length
* 174664 174763: gap of 100 bp
* 174764 176271: contrig of 1508 bp in length
* 176272 176371: gap of 100 bp
* 176372 178022: contrig of 1651 bp in length.
Location/Qualifiers
1. .178022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-683J11"
1. .24283
/note="assembly-fragment"
24384. .46875
/note="assembly-fragment"
46976. .65058
/note="assembly-fragment"
65159. .81322
/note="assembly-fragment"
81423. .98823
/note="assembly-fragment"
98924. .110068
/note="assembly-fragment"
110169. .122716
/note="assembly-fragment"
122817. .133089
/note="assembly-fragment"
133190. .141206
/note="assembly-fragment"
141307. .149050
/note="assembly-fragment"
149151. .156105
/note="assembly-fragment"
156206. .162048
/note="assembly-fragment"
162149. .165118
/note="assembly-fragment clone_end:SP6 vector_side:left"
165219. .167480
/note="assembly-fragment"
167581. .169986
/note="assembly-fragment"
170087. .171909
/note="assembly-fragment"
172010. .172317
/note="assembly-fragment clone_end:T7 vector_side:right"
172418. .174663
/note="assembly-fragment"
174764. .176271
/note="assembly-fragment"
176372. .178022
/note="assembly-fragment"
```

```
BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others
```


ORIGIN

Query Match 34.6%; Score 474; DB 2; Length 178022;

Best Local Similarity 100.0%; Pred No. 1.6e-99;

Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAGACGCCAGCTCACTCATTCACAGGTGAGAGAACATCACCACATATCTGCAA 60
 DB 13428 ATGAGACGCCAGCTCACTCATTCACAGGTGAGAGAACATCACCACATATCTGCAA 13487
 OY 61 GCCAAGCAGCAGACCTGGAAGACCTGCGAGACTTACACAAGATGACAGATAGACA 120
 DB 13488 GCCAAGCAGCAGACCTGGAAGACCTGCGAGACTTACACAAGATGACAGATAGACA 13547
 OY 121 GCCATCAAGTTCACCAACCTGGAGAACGCTTCCAGCTCTTGAGACGGATATGGAAC 180
 DB 13548 GCCATCAAGTTCACCAACCTGGAGAACGCTTCCAGCTCTTGAGACGGATATGGAAC 13607
 OY 181 ATCATTAGCAATATCAGTTACACAGCCCAACACCTGCGAGCTGACCAACAATCTAAT 240
 DB 13608 ATCATTAGCAATATCAGTTACACAGCCCAACACCTGCGAGCTGACCAACAATCTAAT 13667
 OY 241 GAAGTACGAGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 300
 DB 13668 GAAGTACGAGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 13727
 OY 301 AATAATACCTGGCCCAACATCCGTTGGATCTGTTCTCTCAGATGCAACAAGATTG 360
 DB 13728 AATAATACCTGGCCCAACATCCGTTGGATCTGTTCTCTCAGATGCAACAAGATTG 13787
 OY 361 ATGAGCTCAGAGTTAGACACTGAGTAGCCCACTATCACTGATGATGGAAGAATGAG 420
 DB 13788 ATGAGCTCAGAGTTAGACACTGAGTAGCCCACTATCACTGATGATGGAAGAATGAG 13847
 OY 421 CTAGTGACTCCAGCATGCTGCTCATCAACAATTTTCAATFACTACAGGT 474
 DB 13848 CTAGTGACTCCAGCATGCTGCTCATCAACAATTTTCAATFACTACAGGT 13901

RESULT 14

AP001022

LOCUS

AP001022 187635 bp DNA linear HTG 30-MAY-2000

DEFINITION Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING

ACCESSION

AP001022

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 187635)
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Katsushika Univ., 1-15-1 Katsushika, Sagamihara, Kanagawa 228-8555, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923, Fax: 81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:5997772.
 Center: RIKEN Genomic Sciences Center (GSC)
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 Project Information

Center project name: HumDrafl18
 Center clone name: RP11-815L4

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 167192 bases at least Q40

Consensus quality: 177020 bases at least Q30

Consensus quality: 182105 bases at least Q20

Insert size: 184435; sum-of-contigs

Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 21332 contig of 21332 bp in length
 21433 42072 contig of 20640 bp in length
 42173 54336 contig of 12164 bp in length
 54437 66766 contig of 12330 bp in length
 66867 77849 contig of 10983 bp in length
 77950 87798 contig of 9849 bp in length
 87899 99510 contig of 11612 bp in length
 99611 107273 contig of 7663 bp in length
 107374 113189 contig of 5816 bp in length
 113290 120559 contig of 7270 bp in length
 120660 126468 contig of 5809 bp in length
 126569 131619 contig of 5051 bp in length
 131720 136810 contig of 5091 bp in length
 136911 139925 contig of 3015 bp in length
 140026 144232 contig of 4207 bp in length
 144333 148502 contig of 4570 bp in length
 149003 152918 contig of 3916 bp in length
 153019 156399 contig of 3381 bp in length
 156500 158933 contig of 2434 bp in length
 159034 161884 contig of 2851 bp in length
 161985 164301 contig of 2317 bp in length
 164402 166300 contig of 1899 bp in length
 166401 168580 contig of 2180 bp in length
 168681 171270 contig of 2590 bp in length
 171371 174092 contig of 2722 bp in length
 174193 176095 contig of 1903 bp in length
 176196 177890 contig of 1695 bp in length
 177991 179786 contig of 1796 bp in length
 179887 181478 contig of 1592 bp in length
 181579 182895 contig of 1317 bp in length
 182996 184841 contig of 1846 bp in length
 184942 186345 contig of 1404 bp in length
 186446 187635 contig of 1190 bp in length

Sequence updated (26-May-2000).
 NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 21332 contig of 21332 bp in length
 21333 21432 gap of 100 bp
 21433 42072 contig of 20640 bp in length
 42073 42172 gap of 100 bp
 42173 54336 contig of 12164 bp in length
 54337 54436 gap of 100 bp
 54437 66766 contig of 12330 bp in length
 66767 66866 gap of 100 bp
 66867 77849 contig of 10983 bp in length
 77850 77949 gap of 100 bp
 77950 87798 contig of 9849 bp in length
 87799 87898 gap of 100 bp

```
* 87899 99510: contig of 11612 bp in length
* 99511 99610: gap of 100 bp
* 99611 107273: contig of 7663 bp in length
* 107274 107373: gap of 100 bp
* 107374 113189: contig of 5816 bp in length
* 113190 113289: gap of 100 bp
* 113290 120559: contig of 7270 bp in length
* 120560 120659: gap of 100 bp
* 120660 126468: contig of 5809 bp in length
* 126469 126568: gap of 100 bp
* 126569 131619: contig of 5051 bp in length
* 131620 131719: gap of 100 bp
* 131720 136810: contig of 5091 bp in length
* 136811 136910: gap of 100 bp
* 136911 139925: contig of 3015 bp in length
* 139926 140025: gap of 100 bp
* 140026 144232: contig of 4207 bp in length
* 144233 144332: gap of 100 bp
* 144333 148902: contig of 4570 bp in length
* 148903 149002: gap of 100 bp
* 149003 152918: contig of 3916 bp in length
* 152919 153018: gap of 100 bp
* 153019 156399: contig of 3381 bp in length
* 156400 156499: gap of 100 bp
* 156500 158933: contig of 2434 bp in length
* 158934 159033: gap of 100 bp
* 159034 161884: contig of 2851 bp in length
* 161885 161984: gap of 100 bp
* 161985 164301: contig of 2317 bp in length
* 164302 164401: gap of 100 bp
* 164402 166300: contig of 1899 bp in length
* 166301 166400: gap of 100 bp
* 166401 168580: contig of 2180 bp in length
* 168581 168680: gap of 100 bp
* 168681 171270: contig of 2590 bp in length
* 171271 171370: gap of 100 bp
* 171371 174092: contig of 2722 bp in length
* 174093 174192: gap of 100 bp
* 174193 176095: contig of 1903 bp in length
* 176096 176195: gap of 100 bp
* 176196 177890: contig of 1695 bp in length
* 177891 177980: gap of 100 bp
* 177981 179786: contig of 1796 bp in length
* 179787 179886: gap of 100 bp
* 179887 181478: contig of 1592 bp in length
* 181479 181578: gap of 100 bp
* 181579 182895: contig of 1317 bp in length
* 182896 182995: gap of 100 bp
* 182996 184841: contig of 1846 bp in length
* 184842 184941: gap of 100 bp
* 184942 186345: contig of 1404 bp in length
* 186346 186445: gap of 100 bp
* 186446 187635: contig of 1190 bp in length.
*
FEATURES
source
1.187635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-815L4"
1.21332
/misc_feature
/note="assembly_fragment"
21433.42072
/misc_feature
/note="assembly_fragment"
42173.54336
/misc_feature
/note="assembly_fragment"
54437.66766
/misc_feature
/note="assembly_fragment"
66867.77849
/misc_feature
/note="assembly_fragment"
77950.87798
/misc_feature
/note="assembly_fragment clone_end:17 vector_side:right"
87899.99510
```

```
/note="assembly_fragment"
99611.107273
/note="assembly_fragment"
107374.113189
/note="assembly_fragment clone_end:SP6 vector_side:left"
113290.120559
/note="assembly_fragment"
120660.126468
/note="assembly_fragment"
126569.131619
/note="assembly_fragment"
131720.136810
/note="assembly_fragment"
136911.139925
/note="assembly_fragment"
140026.144232
/note="assembly_fragment"
144333.148902
/note="assembly_fragment"
149003.152918
/note="assembly_fragment"
153019.156399
misc_feature
```

Query Match 34.6%; Score 474; DB 2; Length 187635;

Best Local Similarity 100.0%; Pred. No. 1,6e-99;

Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGAACAGCCGAGCTCACTCATTCACAGGTCAGATGGAACATCAGATCTCA 60
Db 58678 ATGAACAGCCGAGCTCACTCATTCACAGGTCAGATGGAACATCAGATCTCA 58737
QY 61 GCCAAGAGAGCAAGACCTGAAAGACCTCGAGACTTACACAAAGATGAGAAATGAGACA 120
Db 58738 GCCAAGAGAGCAAGACCTGAAAGACCTCGAGACTTACACAAAGATGAGAAATGAGACA 58797
QY 121 GCCATCAAGTTCACCAACTGGAGAACGCTTCAGCTCTTTGAGAGGATTTGTGAC 180
Db 58798 GCCATCAAGTTCACCAACTGGAGAACGCTTCAGCTCTTTGAGAGGATTTGTGAC 58857
QY 181 ATCATTTGCAATATCATATTACACAGCCACAGCTGGAGAGCTGACAGCAATCTAAAT 240
Db 58858 ATCATTTGCAATATCATATTACACAGCCACAGCTGGAGAGCTGACAGCAATCTAAAT 58917
QY 241 GAAGTCAGAGCACTTGACAGATACCTTACCAACACAGATGATCTGACCTCTTG 300
Db 58918 GAAGTCAGAGCACTTGACAGATACCTTACCAACACAGATGATCTGACCTCTTG 58977
QY 301 AATATATCCCTGGCCACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATTG 360
Db 58978 AATATATCCCTGGCCACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATTG 59037
QY 361 ATGAGTCGAGGTTAGACACTGAAGTGAAGCAACTTATCAGTATGATGGAAGAAATGAG 420
Db 59038 ATGAGTCGAGGTTAGACACTGAAGTGAAGCAACTTATCAGTATGATGGAAGAAATGAG 59097
QY 421 CTACTAGACTGCAAGCATGCTCAGCTCATCAAGATTTTACAAATCTCAAGCT 474
Db 59098 CTACTAGACTGCAAGCATGCTCAGCTCATCAAGATTTTACAAATCTCAAGCT 59151
RESULT 15
AP000939 188255 bp DNA linear HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING
DEFINITION DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP000939.3 GI:8119080
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens DNA, Clone:RP11-839023.
SOURCE Homo sapiens
ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188255)
```

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Homo sapiens 188,255 genomic DNA of 18p11.3
 JOURNAL Published Only in Database (1999)
 REFERENCE 2 (bases 1 to 188255)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 Kitasato Univ., 1-15-1 Kitasato, Sagaminara, Kanagawa 228-8555,
 Japan (E-mail:hattori@gsc.riken.go.jp,
 URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
 Fax:81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:6997751.

COMMENT

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

Project Information

Center project name: Humdraft18

Center clone name: RPI1-839023

Summary Statistics

Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329

Consensus quality: 171068 bases at least Q40

Consensus quality: 182445 bases at least Q20

Insert size: 185155; sum-of-contigs

Quality coverage: 4.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
 32 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs 'N', but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

```

1      16096 contig of 16096 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 63558 contig of 12448 bp in length
63659 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 6072 bp in length
91010 98615 contig of 7606 bp in length
98716 105978 contig of 7263 bp in length
106079 113049 contig of 6971 bp in length
113150 119201 contig of 6052 bp in length
119302 126582 contig of 7281 bp in length
126683 132070 contig of 5388 bp in length
132171 137138 contig of 4968 bp in length
137239 143055 contig of 5817 bp in length
143156 148639 contig of 5484 bp in length
148740 153086 contig of 4347 bp in length
153187 158922 contig of 5735 bp in length
159022 163174 contig of 4133 bp in length
163275 167066 contig of 3792 bp in length
167167 170120 contig of 2954 bp in length
170221 172430 contig of 2210 bp in length
172531 175598 contig of 3068 bp in length
175699 177118 contig of 1420 bp in length
177219 178531 contig of 1313 bp in length
178632 179857 contig of 1107 bp in length
179958 181064 contig of 1571 bp in length
181165 182735 contig of 1088 bp in length
182836 183923 contig of 1787 bp in length
184024 185810 contig of 1200 bp in length
185911 187145 contig of 1235 bp in length
187246 188255 contig of 1010 bp in length
Sequence updated (06-Jan-2000)

```

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of 'N', but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      16096: contig of 16096 bp in length
16097 16196: gap of 100 bp
16197 33789: contig of 17593 bp in length
33790 33889: gap of 100 bp
33890 51010: contig of 17121 bp in length
51011 51110: gap of 100 bp
51111 51110: contig of 12448 bp in length
51111 63558: gap of 100 bp
63559 63658: gap of 100 bp
63659 75056: contig of 11398 bp in length
75057 75156: gap of 100 bp
75157 84737: contig of 9581 bp in length
84738 84837: gap of 100 bp
84838 90909: contig of 6072 bp in length
90910 91009: gap of 100 bp
91010 98615: contig of 7606 bp in length
98616 98715: gap of 100 bp
98716 105978: contig of 7263 bp in length
105979 106078: gap of 100 bp
106079 113049: contig of 6971 bp in length
113050 113149: gap of 100 bp
113150 119201: contig of 6052 bp in length
119202 119301: gap of 100 bp
119302 126582: contig of 7281 bp in length
126583 126682: gap of 100 bp
126683 132070: contig of 5388 bp in length
132071 132170: gap of 100 bp
132171 137138: contig of 4968 bp in length
137139 137238: gap of 100 bp
137239 143055: contig of 5817 bp in length
143056 143155: gap of 100 bp
143156 148639: contig of 5484 bp in length
148640 148739: gap of 100 bp
148740 153086: contig of 4347 bp in length
153087 153186: gap of 100 bp
153187 158921: contig of 5735 bp in length
158922 159021: gap of 100 bp
159022 163174: contig of 4133 bp in length
163175 163274: gap of 100 bp
163275 167066: contig of 3792 bp in length
167067 167166: gap of 100 bp
167167 170120: contig of 2954 bp in length
170121 170220: gap of 100 bp
170221 172430: contig of 2210 bp in length
172431 172530: gap of 100 bp
172531 175598: contig of 3068 bp in length
175599 175698: gap of 100 bp
175699 177118: contig of 1420 bp in length
177119 177218: gap of 100 bp
177219 178531: contig of 1313 bp in length
178532 178631: gap of 100 bp
178632 179857: contig of 1107 bp in length
179858 179957: gap of 100 bp
179958 181064: contig of 1571 bp in length
181065 181164: gap of 100 bp
181165 182735: contig of 1088 bp in length
182736 182835: gap of 100 bp
182836 183923: contig of 1787 bp in length
183924 184023: gap of 100 bp
184024 185810: contig of 1200 bp in length
185811 185910: gap of 100 bp
185911 187145: contig of 1235 bp in length
187146 187245: gap of 100 bp
187246 188255: contig of 1010 bp in length.
Location/Qualifiers

```

FEATURES

```
source
1. .188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1. .16096
/note="assembly_fragment"
16197. .33789
/note="assembly_fragment"
33890. .51010
/note="assembly_fragment"
51111. .63538
/note="assembly_fragment"
63659. .75056
/note="assembly_fragment"
75157. .84737
/note="assembly_fragment"
84838. .90909
/note="assembly_fragment"
91010. .98615
/note="assembly_fragment"
98716. .105978
/note="assembly_fragment"
106079. .113049
/note="assembly_fragment"
113150. .119201
/note="assembly_fragment"
119302. .126582
/note="assembly_fragment"
126683. .132070
/note="assembly_fragment"
132171. .137138
/note="assembly_fragment"
137239. .143055
/note="assembly_fragment"
143156. .148639
/note="assembly_fragment"
148740. .153086
/note="assembly_fragment"
153187. .158921
/note="assembly_fragment"
159022. .163174
misc_feature
33.7%; Score 462; DB 2; Length 188255;
Best Local Similarity 99.8%; Pred. No. 1e-96;
Matches 473; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 ATGAACAGCCAGCTCACTATTCACAGGTCAAGATGAGAAACATCACCCTATCTCTCAA 60
Db 102931 ATGAACAGCCAGCTCACTATTCACAGGTCAAGATGAGAAACATCACCCTATCTCTCAA 102990
QY 61 GCCAACGAGCAGAACTGGAAGACCTGACAGACTTACACAAAGATGCAGAAATAGAAC 120
Db 102991 GCCAACGAGCAGAACTGGAAGACCTGACAGACTTACACAAAGATGCAGAAATAGAAC 103050
QY 121 GCCATCAAGTTCAACCACTGAGAGAAAGCTTCCAGCTTTTGAGACGATATTGTGAAC 180
Db 103051 GCCATCAAGTTCAACCACTGAGAGAAAGCTTCCAGCTTTTGAGACGATATTGTGAAC 103110
QY 181 ATCATTTAGCAATATCAGTTACACAGCCACACCTGCGAGCGTGACACCAATCTAAAT 240
Db 103111 ATCATTTAGCAATATCAGTTACACAGCCACACCTGCGAGCGTGACACCAATCTAAAT 103170
QY 241 GAAGTCAGAGCACTGCAAGATACCTTACAAACACAGATGATGACCTCCTTG 300
Db 103171 GAAGTCAGAGCACTGCAAGATACCTTACAAACACAGATGATGACCTCCTTG 103230
QY 301 AATAATACCTGGCCAAACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATTG 360
Db 103231 AATAATACCTGGCCAAACATCCGTTT-GATTCTGTTCTCTCAGATGCAACAAGATTG 103289
QY 361 ATGAGGTGCGAGTTAGACACTGAAGTACCACTTATCAGTGAATTATGGAAGAAATGAAG 420
```

```
|||||
Db 103290 ATGAGGTGCGAGTTAGACACTGAAGTACGCCACTTATCAGTGAATTGGAAGAAATGAAG 103349
QY 421 CTAGTACTCCAAAGCATGGTCAAGTCAATCAAGATTTTACATTAATCTACAGGT 474
|||||
Db 103350 CTAGTACTCCAAAGCATGGTCAAGTCAATCAAGATTTTACATTAATCTACAGGT 103403
```

Search completed: March 21, 2003, 06:35:22
Job time : 4123.17 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 272.589 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712a-1_COPY_325_1695
Perfect score: 1371
Sequence: 1 atgacagccagctcaactc.....cagctactgcatcgtcattca 1371

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.GeneSeq_101002:*
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT:*
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT:*
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT:*
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT:*
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT:*
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT:*
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT:*
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT:*
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT:*
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT:*
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT:*
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT:*
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT:*
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT:*
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT:*
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT:*
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT:*
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT:*
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT:*
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT:*
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT:*
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2001.DAT:*
24: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1371	100.0	2024	AAA07697	Human collectin en
2	1371	100.0	2262	ABA97932	Human scavenger re
3	1371	100.0	2628	AAH43036	Nucleotide sequenc
4	1369.4	99.9	2181	AAI60628	Human polynucleoti
5	1369.4	99.9	2641	AAQ66903	Human EXMAD-14 cod
6	1369.4	99.9	2929	ABQ92072	Human polynucleoti
7	1369.4	99.9	2930	AAV55746	Human secreted pro
8	1358.4	99.1	2318	AAI58842	Human polynucleoti
9	1357.4	99.0	2005	ABL95574	Human angiogenesis

10	1357.4	99.0	2005	24	ABL88085	Human PRO7223 CDNA
11	1086.2	79.2	2637	22	AAH43037	Nucleotide sequenc
12	883.8	64.5	1521	23	AAI71133	DNA encoding novel
13	617	45.0	2256	22	AAH43054	Nucleotide sequenc
14	235.6	17.2	3685	20	AAI27858	Human CSR2 protein
15	235.6	17.2	3810	20	AAI27856	Human CSR1 protein
16	169.4	12.4	493	23	AAI71131	DNA encoding novel
17	150.2	11.0	873	23	AAI71130	DNA encoding novel
18	150.2	11.0	1062	23	AAI67442	DNA encoding novel
19	146.8	10.7	1329	23	AAI71134	DNA encoding novel
20	143	10.4	5041	22	AAI17153	Human nervous syst
21	123	9.0	1877	20	AAI27857	Human CSR2 protein
22	121.6	8.9	756	14	AAQ43034	Collagen-like poly
23	121.6	8.9	5756	17	AAI16768	Collagen-like poly
24	121.4	8.9	5676	15	AAQ64556	Human collagen (TY
25	121.4	8.9	8284	22	AAI26526	Human breast cance
26	121.4	8.9	8284	22	AAI26553	Human breast cance
27	121.4	8.9	8284	22	AAI26600	Human breast cance
28	121	8.8	1707	22	AAI44989	CDNA encoding nove
29	118.4	8.6	6109	23	AAI78667	Murine Col5a3 CDNA
30	118	8.6	756	14	AAQ43032	Collagen-like poly
31	118	8.6	756	17	AAI16766	Collagen-like poly
32	117	8.5	9287	24	AAI64501	Human benign prost
33	115.8	8.4	4748	22	AAI06573	Human alpha1(I) c
34	115.2	8.4	986	22	AAI89468	Human polynucleoti
35	115	8.4	4270	24	AAI99885	Mouse ischaemic co
36	114.8	8.4	4816	24	AAI84041	Human CDNA differe
37	114.8	8.4	5058	23	AAI63026	DNA encoding novel
38	114.8	8.4	6358	22	AAI98286	Human EST-derived
39	114.8	8.4	6691	23	AAI79806	DNA encoding novel
40	114	8.3	5060	24	AAI92176	Prostate cancer-as
41	113.8	8.3	506	17	AAI12198	pUG4-5-CDK-BP CDNA
42	113.6	8.3	3181	19	AAV53358	Nucleotide sequenc
43	113.4	8.3	4821	23	AAI86866	DNA encoding novel
44	113.4	8.3	6200	23	AAI78668	Human COL5A3 CDNA
45	113.4	8.3	6368	24	AAI59647	Novel human coding

ALIGNMENTS

RESULT 1	AAA07697	standard; DNA; 2024 BP.
ID	AAA07697	
AC	AAA07697;	
XX		
DT	20-JUN-2000	(first entry)
XX		
DE	Human collectin encoding DNA.	
XX		
KW	Collectin; human; antibacterial; antiviral; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	55..1698
FT		/*tag= "a"
FT		/product= "collectin"
FT	misc_feature	1..738
FT		/*tag= "b"
FT		/note= "fragment specifically claimed in claim 8"
FT	misc_feature	55..738
FT		/*tag= "c"
FT		/note= "fragment specifically claimed in claim 8"
FT	misc_feature	79..738
FT		/*tag= "d"
FT		/note= "fragment specifically claimed in claim 8"
FT	misc_feature	325..738
FT		/*tag= "e"
FT		/note= "fragment specifically claimed in claim 8"
FT	misc_feature	358..738
FT		/*tag= "f"

FT	/note= "fragment specifically claimed in claim 8"
FT	misc_feature
FT	670..1695
FT	/*tag= g
FT	/note= "fragment specifically claimed in claim 5"
FT	misc_feature
FT	685..738
FT	/*tag= h
FT	/note= "fragment specifically claimed in claim 7"
FT	misc_feature
FT	730..738
FT	/*tag= i
FT	/note= "fragment specifically claimed in claim 7"
FT	misc_feature
FT	739..1695
FT	/*tag= j
FT	/note= "fragment specifically claimed in claim 6"
FT	misc_feature
FT	1696..2024
FT	/*tag= k
FT	/note= "fragment specifically claimed in claim 9"
PN	WO200011161-A1.
XX	
PD	02-MAR-2000.
XX	
PE	24-AUG-1999; 99WO-JP04552.
XX	
PR	24-AUG-1998; 98JP-0237611.
XX	
PA	(FUSO) FUSO PHARM IND LTD.
XX	
PI	Wakamiya N;
XX	
DR	WPI: 2000-224696/19.
XX	
PT	P-PSDB: AA177985.
XX	
PT	New collectin of human origin having antibacterial and antiviral
PT	activity, and gene encoding it useful for production of transgenic
PT	animals and of antibodies for screening potential drug molecules -
XX	
PS	Claim 5-9; Page 87-90; 106pp; Japanese.
XX	
CC	The invention relates to polynucleotides encoding a new collectin of
CC	human origin. The collectin can be used as an antibacterial and antiviral
CC	agent and for screening potential drug molecules. The new collectin can
CC	be produced by standard recombinant methodology. The present sequence
CC	represents a DNA encoding the human collectin.
XX	
XX	
XX	Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;

	Query Match	Similarity	100.0%	Score 1371:	DB 21:	Length 204:	
	Best Local	Similarity	100.0%	Pred. No. 0:			
	Matches 1371:	Conservative	0:	Mismatches	0:	Indels	Gaps
OY	1	ATGAACAGCCACCTCAACATTCATTCACAGGTGAGATGAGAACATCAACCTATCTCTCAA	60				
Db	325	ATGAACAGCCAGCCTCAACATTCATTCACAGGTGAGATGAGAACATCAACCTATCTCTCAA	384				
OY	61	GCCAAAGCAGCAACCTGAAAGACCTGCAGACTTACACAAAGATGCAGAAATAGACA	120				
Db	385	GCCAAAGCAGCAACCTGAAAGACCTGCAGACTTACACAAAGATGCAGAAATAGACA	444				
OY	121	GCCATCAGTTCACCAACATGAGAGAACGCTTCACGCTCTTTGAGAGGATATTGTGAAC	180				
Db	445	GCCATCAGTTCACCAACATGAGAGAACGCTTCACGCTCTTTGAGAGGATATTGTGAAC	504				
OY	181	ATCATTTAGCAATATATCAGTTTACACAGCCCAACACATGCGGAGCGCTGCACGCAATCTTAAT	240				
Db	505	ATCATTTAGCAATATATCAGTTTACACAGCCCAACACATGCGGAGCGCTGCACGCAATCTTAAT	564				
OY	241	GAAGTCAGGACCACTTGACACAGATACCTTTACCAAAACACACAGATGATTCGACCTCTTG	300				
Db	565	GAAGTCAGGACCACTTGACACAGATACCTTTACCAAAACACACAGATGATTCGACCTCTTG	624				
OY	301	AATATATACCCCTGCGCAACATCCGTTTGATTCGTTTCTCTAGAGATGCAACAAGATTTG	360				
Db	625	AATATATACCCCTGCGCAACATCCGTTTGATTCGTTTCTCTAGAGATGCAACAAGATTTG	684				

Qy	361	ATGAGCTCGAGGTTTAGCACTGGAAGTAGCCAACTTATTCAGTGAATGTTATGGAAGAAATGAG	420
Db	685	ATGAGCTCGAGGTTTAGCACTGGAAGTAGCCAACTTATTCAGTGAATGTTATGGAAGAAATGAG	744
Qy	421	CTAGTAGACTCCAAAGCATGTGTACGCTCATCCAGAAATTTTACATACTACAAAGTCTCACCG	480
Db	745	CTAGTAGACTCCAAAGCATGTGTACGCTCATCCAGAAATTTTACATACTACAAAGTCTCACCG	804
Qy	481	GGCCCCAGGGGTCCAGAGGTGACAGAGATATCCAGGACCCCTGTGGCCAACTGGCAAC	540
Db	805	GGCCCCAGGGGTCCAAAGGTGACAGAGATATCCAGGACCCCTGTGGCCAACTGGCAAC	864
Qy	541	AAGGACAGAAAGAGAGAGAGGGGAGACCCGTGACACCTCTGGCCCTCGGGGTAGAGAGAC	600
Db	865	AAGGACAGAAAGAGAGAGAGGGGAGACCCGTGACACCTCTGGCCCTCGGGGTAGAGAGAC	924
Qy	601	CCAATTGGACCAAGCTGGTCCCTCCCGAGAGACGTGGCGGCAAAAGATCTTAAAGCTCCAG	660
Db	925	CCAATTGGACCAAGCTGGTCCCTCCCGAGAGACGTGGCGGCAAAAGATCTTAAAGCTCCAG	984
Qy	661	GGCCCCAAAGGCTCCGGTGTCCCTCGGGAGACCCGGGCTCAGAGGCCCGAGTGGGAC	720
Db	985	GGCCCCAAAGGCTCCGGTGTCCCTCGGGAGACCCGGGCTCAGAGGCCCGAGTGGGAC	1044
Qy	721	CCAGAGCCCCCGGGCCCAACAGGCAAAAGAGGACTCCCGGGCCCTCAGAGGCCCTCTGGC	780
Db	1045	CCAGAGCCCCCGGGCCCAACAGGCAAAAGAGGACTCCCGGGCCCTCAGAGGCCCTCTGGC	1104
Qy	781	TTCCAGGAGACTTCAGGGCACCCGTTGGGAGACCTGGGGTCTGTGACCTCGGGGACTGCCA	840
Db	1105	TTCCAGGAGACTTCAGGGCACCCGTTGGGAGACCTGGGGTCTGTGACCTCGGGGACTGCCA	1164
Qy	841	GGCTTGCGCTGGGGATCCAGGCAATGAGCCAGAGGCCCAAGGGCCCCCGGGCTCTGGGCCA	900
Db	1165	GGCTTGCGCTGGGGATCCAGGCAATGAGCCAGAGGCCCAAGGGCCCCCGGGCTCTGGGCCA	1224
Qy	901	TCAGAGACGCGTGGTGGCCCTCGGCCCTTGACAGAAATGAGCCAAACCCCGGACACGGAGACAAT	960
Db	1225	TCAGAGACGCGTGGTGGCCCTCGGCCCTTGACAGAAATGAGCCAAACCCCGGACACGGAGACAAT	1284
Qy	961	GGCTGGCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA	1020
Db	1285	GGCTGGCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA	1344
Qy	1021	GAAATTTTGTAGAGATGCAAGGCTTTCTGTGGAAGCAACATCTTCACATCTGTTTCATA	1080
Db	1345	GAAATTTTGTAGAGATGCAAGGCTTTCTGTGGAAGCAACATCTTCACATCTGTTTCATA	1404
Qy	1081	AACACTGAGAGAGAAACAGCAATGATAAAAAAAACAGATGTTAGGAGAGAGACCACTGG	1140
Db	1405	AACACTGAGAGAGAAACAGCAATGATAAAAAAAACAGATGTTAGGAGAGAGACCACTGG	1464
Qy	1141	ATCGGCGCTCACAGACTCAGAGCGTGAAATGAATGAGTAAGTGGCTGTGATGGCACTTCCA	1200
Db	1465	ATCGGCGCTCACAGACTCAGAGCGTGAAATGAATGAGTAAGTGGCTGTGATGGCACTTCCA	1524
Qy	1201	GACTACAAAAAATTTGGAAAGCTGGACAGCCGGATAACTGGGGTCAATGACCATGGGCGCAGGA	1260
Db	1525	GACTACAAAAAATTTGGAAAGCTGGACAGCCGGATAACTGGGGTCAATGACCATGGGCGCAGGA	1584
Qy	1261	GAAGACTGTGCTGGGTGATTTATGCTGGGCACTGGAAACGATTTCCAAATGTGAAACACTC	1320
Db	1585	GAAGACTGTGCTGGGTGATTTATGCTGGGCACTGGAAACGATTTCCAAATGTGAAACACTC	1644
Qy	1321	AATTACTTCAATTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1371	
Db	1645	AATTACTTCAATTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1695	
RESULT 2			
AB97932 standard; cDNA: 2262 BP			

```
XX ABA97932;
AC
XX
XX 25-APR-2002 (first entry)
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
XX Human scavenger receptor; diagnosis; treatment; autoimmune disease;
XX rheumatoid arthritis; gene; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 33..2262
XX FT /*tag=a
XX FT /product="scavenger receptor-like protein"
XX
XX JP2001340089-A.
XX
XX 11-DEC-2001.
XX
XX 08-DEC-2000; 2000JP-0375066.
XX
XX 27-MAR-2000; 2000JP-0090772.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX WPI: 2002-144965/19.
XX P-PSDB; ABB08642.
XX
XX New scavenger receptor-like protein for diagnosis, prevention and
XX treatment of autoimmune disease, such as rheumatoid arthritis -
XX
XX Claim 3; Fig 1; 38pp; Japanese.
XX
XX The invention relates to a human scavenger receptor-like protein. The
XX protein is useful as a target molecule for diagnosis, prevention and
XX treatment of autoimmune diseases such as rheumatoid arthritis.
XX
XX Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other:
XX
Query Match 100.0%; Score 1371; DB 24; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGAACAGCCAGCTCACTCACTTACAGAGTCAGATGAGAACATCACTATCTCTCAA 60
DB 889 ATGAACAGCCAGCTCACTCACTTACAGAGTCAGATGAGAACATCACTATCTCTCAA 948
QY 61 GCCAAGAGCAGACCTGAAAGACCTGCAGACTTACAAAGATGACAGAAATAGAAC 120
DB 949 GCCAAGAGCAGACCTGAAAGACCTGCAGACTTACAAAGATGACAGAAATAGAAC 1008
QY 121 GCCATCAAGTTAACCACTGAGAGAACGCTTCCAGCTCTTGTGAGACGATATTGTGAAC 180
DB 1009 GCCATCAAGTTAACCACTGAGAGAACGCTTCCAGCTCTTGTGAGACGATATTGTGAAC 1068
QY 181 ATCATTAAGCAATATGATGATACAGACGCCACGCTGCGAGCCTGACCGCATCTAAAT 240
DB 1069 ATCATTAAGCAATATGATGATACAGACGCCACGCTGCGAGCCTGACCGCATCTAAAT 1128
QY 241 GAAGTCAGAGCAGCTTGCACAGATACCTTACCAAAACAGACAGATATGACCTCTTG 300
DB 1129 GAAGTCAGAGCAGCTTGCACAGATACCTTACCAAAACAGACAGATATGACCTCTTG 1188
QY 301 AATTAATACCTTGCCCAACATCCGTTTGATTTCTGTTCTCTCAGATGCAACAAGATTG 360
DB 1189 AATTAATACCTTGCCCAACATCCGTTTGATTTCTGTTCTCTCAGATGCAACAAGATTG 1248
QY 361 ATGAGGTGAGGTGAGACCTGAGTAGCCACTTATCTGATTTATGGAAGAAATGAAG 420
DB 1249 ATGAGGTGAGGTGAGACCTGAGTAGCCACTTATCTGATTTATGGAAGAAATGAAG 1308
```

```
QY 421 CTAGTAGACTCCAGACATGATGATCATCATCAATTTTACAAATTTACTACAGATCCACCG 480
DB 1309 CTAGTAGACTCCAGACATGATGATCATCATCAATTTTACTACAGATCCACCG 1368
QY 481 GGGCCCCAGGGGTTCCAGAGGTGACAGAGATCCAGAGACCCCTGGCCCACTGGCAAC 540
DB 1369 GGGCCCCAGGGGTTCCAGAGGTGACAGAGATCCAGAGACCCCTGGCCCACTGGCAAC 1428
QY 541 AAGGACACAAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCTGGGTGAGAGAGC 600
DB 1429 AAGGACACAAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCTGGGTGAGAGAGC 1488
QY 601 CCAATTGAGACAGACTGATCCCGCCGAGAGACGGTGGGAGCAAAAGATCTAAAGCTCCAG 660
DB 1489 CCAATTGAGACAGACTGATCCCGCCGAGAGACGGTGGGAGCAAAAGATCTAAAGCTCCAG 1548
QY 661 GGGCCCCAAGGCTCCCGTGGTCCCTGGGAAAGCCCGGCCCTCAGAGGCCCACTGGGAGC 720
DB 1549 GGGCCCCAAGGCTCCCGTGGTCCCTGGGAAAGCCCGGCCCTCAGAGGCCCACTGGGAGC 1608
QY 721 CCAGGCCCCCGGGCCCAACAGAGGACTCCCGGCCCTTACAGGACCTCCTGGC 780
DB 1609 CCAGGCCCCCGGGCCCAACAGAGGACTCCCGGCCCTTACAGGACCTCCTGGC 1668
QY 781 TTCCAGGAGCTTCAGGAGCAACCGTTGGGAGACCTGGGGTGCCTGAGCTGGGAGCTGCA 840
DB 1669 TTCCAGGAGCTTCAGGAGCAACCGTTGGGAGACCTGGGGTGCCTGAGCTGGGAGCTGCA 1728
QY 841 GGCCTTCCTGGGGTACAGGACATGCCAGGCCCAAGGGGCCCGGCCCTCTCGGCCCA 900
DB 1729 GGCCTTCCTGGGGTACAGGACATGCCAGGCCCAAGGGGCCCGGCCCTCTCGGCCCA 1788
QY 901 TCAGAGAGCGGTGTGTCCTGCGCTGCGAGAAATGACACCAACCCCGGACCGGAGACAT 960
DB 1789 TCAGAGAGCGGTGTGTCCTGCGCTGCGAGAAATGACACCAACCCCGGACCGGAGACAT 1848
QY 961 GCGTGCCTCCCTACACAGAAACTTCACAGACAAATGCTACTATTTTCAGTGAAGAA 1020
DB 1849 GCGTGCCTCCCTACACAGAAACTTCACAGACAAATGCTACTATTTTCAGTGAAGAA 1908
QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGTAAGACAGCTCTTCATCTGTTTCATA 1080
DB 1909 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGTAAGACAGCTCTTCATCTGTTTCATA 1968
QY 1081 AACACTAGAGAGAGAACAGCAATGATTAATAAAACAGATGATGAGAGAGAGCCACTGG 1140
DB 1969 AACACTAGAGAGAGAACAGCAATGATTAATAAAACAGATGATGAGAGAGAGCCACTGG 2028
QY 1141 ATCGGCTCAGACACTCAGAGCGTGAATAATGAATGAGAGTGGCTGATGGAGCATCTCA 1200
DB 2029 ATCGGCTCAGACACTCAGAGCGTGAATAATGAATGAGAGTGGCTGATGGAGCATCTCA 2088
QY 1201 GACTACAAAATTTGGAAGCTGAGACGCCGATTAATGAGGAGTATGAGCCATGGCCAGCA 1260
DB 2089 GACTACAAAATTTGGAAGCTGAGACGCCGATTAATGAGGAGTATGAGCCATGGCCAGCA 2148
QY 1261 GAAGACTGTGCTGGTGAATTTATGCTGGGCAAGTGAAGATTTCCAAATGTGAAGACGT 1320
DB 2149 GAAGACTGTGCTGGTGAATTTATGCTGGGCAAGTGAAGATTTCCAAATGTGAAGACGT 2208
QY 1321 AATTAATCTATTGGCAAAAAGACAGAGGAGACAGTACTGTCACTCATCTTA 1371
DB 2209 AATTAATCTATTGGCAAAAAGACAGAGGAGACAGTACTGTCACTCATCTTA 2259
RESULT 3
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX
XX AAH43036;
AC
XX
XX 15-OCT-2001 (first entry)
XX
```


DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human: scavenger receptor; SRCL-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
XX low density lipoprotein; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 74..2302
FT /tag= a
FT /product= "scavenger receptor"
XX
PN WO200159107-A1.
PD 16-AUG-2001.
PF 08-FEB-2001: 2001WO-JP00874.
XX
PR 14-FEB-2000: 2000JP-0035155.
PR 10-OCT-2000: 2000JP-0309068.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N:
XX
DR WPI: 2001-497076/54.
DR P-PDB: AAG63346.
XX
PT New Scavenger receptor proteins SRCL-PI with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation
PS Claim 2; Page 79-84; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
CC SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SQ Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;
Query Match 100.0%; Score 1371; DB 22; Length 2628;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAACAGCCAGCTCACTCATTCACAGTCAGATGAGAAATACACACTATCTCTCAA 60
DB 929 ATGAACAGCCAGCTCACTCATTCACAGTCAGATGAGAAATACACACTATCTCTCAA 988
QY 61 GCCAACGACGAGAACTGGAAGACCTGACGACTTACACAAAGATGCGAAGATAGAAC 120
DB 989 GCCAACGACGAGAACTGGAAGACCTGACGACTTACACAAAGATGCGAAGATAGAAC 1048
QY 121 GCCATCAAGTTCAACGACGAGGAAAGCTTCAGCTCTTGAGANGATATTGTGAAC 180
DB 1049 GCCATCAAGTTCAACGACGAGGAAAGCTTCAGCTCTTGAGANGATATTGTGAAC 1108
QY 181 ATCATTTAGCAATATACATTACAGCCACACCTGGGAGCTGACCAATCTTAAT 240
DB 1109 ATCATTTAGCAATATACATTACAGCCACACCTGGGAGCTGACCAATCTTAAT 1168
QY 241 GAAGTCAGGACCACTGTCACAGATACCTTACCAAAACACAGATGATGTACCTCTTG 300
DB 1169 GAAGTCAGGACCACTGTCACAGATACCTTACCAAAACACAGATGATGTACCTCTTG 1228
QY 301 AATTAATACCTGGCGAAGATCCGTTGGATTCTCTCTCAGATGCAAGATTTG 360
|||||

DB 1229 AATAATACCTGGCGAAGATCCGTTGGATTCTCTCTCAGATGCAAGATTTG 1288
QY 361 ATGAGTCGAGGTTAGACATGAAATAGCCAACTTACATGATTAATGGAAGAAATGAG 420
DB 1289 ATGAGTCGAGGTTAGACATGAAATAGCCAACTTACATGATTAATGGAAGAAATGAG 1348
QY 421 CTAGTAGACTCCAAAGCATGCTCAGCTCATCAACAATTTTACATACTACAGATCCACG 480
DB 1349 CTAGTAGACTCCAAAGCATGCTCAGCTCATCAACAATTTTACATACTACAGATCCACG 1408
QY 481 GGGCCCAAGGGGTCGAAGAGTGACAGAGATGCCAGGAGCCCTGGCCCAACTGGCAAC 540
DB 1409 GGGCCCAAGGGGTCGAAGAGTGACAGAGATGCCAGGAGCCCTGGCCCAACTGGCAAC 1468
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGCCCTGAGCACTTGCCCTCGGCTGAGAGAGGC 600
DB 1469 AAGGACAGAAAGAGAGAGAGGGGAGCCCTGAGCACTTGCCCTCGGCTGAGAGAGGC 1528
QY 601 CCAATTGACACAGCTGCTCCCTCCCGGAGAGCTGGCGGCAAGAGATCTAAAGCTCCAG 660
DB 1529 CCAATTGACACAGCTGCTCCCTCCCGGAGAGCTGGCGGCAAGAGATCTAAAGCTCCAG 1588
QY 661 GGGCCCAAGGGGTCGCTGCTCCCTCGGAGAGCCCGGCCCTCAGGGGCCCACTGGGGAC 720
DB 1589 GGGCCCAAGGGGTCGCTGCTCCCTCGGAGAGCCCGGCCCTCAGGGGCCCACTGGGGAC 1648
QY 721 CCAAGGCCCCCGGGCCACCAGGCAAAAGAGAGACTCCCGGCCCTCAGGGGCCCTCTGAGC 780
DB 1649 CCAAGGCCCCCGGGCCACCAGGCAAAAGAGAGACTCCCGGCCCTCAGGGGCCCTCTGAGC 1708
QY 781 TTCACAGGACTTCAGGGACACGTTGGGGAGCTGGGGGCTGAGACTTGGGGAGCTGCCA 840
DB 1709 TTCACAGGACTTCAGGGACACGTTGGGGAGCTGGGGGCTGAGACTTGGGGAGCTGCCA 1768
QY 841 GGCCTGCTGGGGTACCAGGATCCAGAGCCCAAGGGCCCGCGGCCCTCTGGGCCCA 900
DB 1769 GGCCTGCTGGGGTACCAGGATCCAGAGCCCAAGGGCCCGCGGCCCTCTGGGCCCA 1828
QY 901 TCAGAGCGGCTGTGCTGCCCTCGGCTGCAAGATGAGCAACCCCGGACCGAGAGCAAT 960
DB 1829 TCAGAGCGGCTGTGCTGCCCTCGGCTGCAAGATGAGCAACCCCGGACCGAGAGCAAT 1888
QY 961 GGCCTGCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
DB 1889 GGCCTGCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1948
QY 1021 GAATTTTGGAGATGCAAAAGCTTTCTGTGAGAGACAAGCTCACATCTGTTTCATA 1080
DB 1949 GAATTTTGGAGATGCAAAAGCTTTCTGTGAGAGACAAGCTTCACATCTGTTTCATA 2008
QY 1081 AACACTAGAGAGAGACAGCAATGATTAATAAAGACAGATGAGAGAGAGACCACTGG 1140
DB 2009 AACACTAGAGAGAGACAGCAATGATTAATAAAGACAGATGAGAGAGAGACCACTGG 2068
QY 1141 ATGCGCTCAGACACTGAGAGCTGAATGAATGAATGAGTGGCTGATGGGACATCTCCA 1200
DB 2069 ATGCGCTCAGACACTGAGAGCTGAATGAATGAATGAGTGGCTGATGGGACATCTCCA 2128
QY 1201 GACTACAAAATTTGGAAGCTGGACAGCCGATTAAGTGGGGTATGGCCATGGGCGCAGGA 1260
DB 2129 GACTACAAAATTTGGAAGCTGGACAGCCGATTAAGTGGGGTATGGCCATGGGCGCAGGA 2188
QY 1261 GAAGACTGTGCTGGGTTGATTTATGTGGGCACTGGAAGCAATTTCCAAATGTGAAGACGT 1320
DB 2189 GAAGACTGTGCTGGGTTGATTTATGTGGGCACTGGAAGCAATTTCCAAATGTGAAGACGT 2248
QY 1321 AATAACTTCAATTTTGGGAAAAAGACAGGAGACAGTACTGTCAATCTCATTTA 1371
DB 2249 AATAACTTCAATTTTGGGAAAAAGACAGGAGACAGTACTGTCAATCTCATTTA 2299
RESULT 4
AAT160628

ID AA160628 standard; cDNA; 2181 BP.
XX
AC AA160628;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4617.
XX
KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokineic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PE 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HXSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
P-PSDB: AAM41472.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4617; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokineic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;

Query Match 99.9%; Score 1369.4; DB 22; Length 2181;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGAACAGCAGCAGCTCACTGATTCACAGGTGAGAGAAATCATACCACTATCTCTCAA 60
DB 317 ATGAACAGCAGCAGCTCACTGATTCACAGGTGAGAGAAATCATACCACTATCTCTCAA 376

OY 61 GCCAAGCAGCAGAACCTGTAAGACCTGAGAGCTTACCAAAAGATCAGAGATAGAAC 120
DB 377 GCCAAGCAGCAGAACCTGTAAGACCTGAGAGCTTACCAAAAGATCAGAGATAGAAC 436
OY 121 GCCATCAAGTTCAACCACTGAGAGAGCGCTTCCAGCTCTTGGAGCGATATTTGGAAC 180
DB 437 GCCATCAAGTTCAACCACTGAGAGAGCGCTTCCAGCTCTTGGAGCGATATTTGGAAC 496
OY 181 ATCATATGCAATATATGCTTACACAGCCACACCGCCAGCGAGCTGACCAATCAAT 240
DB 497 ATCATATGCAATATATGCTTACACAGCCACACCGCCAGCGAGCTGACCAATCAAT 556
OY 241 GAAGTCAGAGACACTTGCACAGATACCTTACCAAAACACAGATATGACCTCTTG 300
DB 557 GAAGTCAGAGACACTTGCACAGATACCTTACCAAAACACAGATATGACCTCTTG 616
OY 301 AATATATACCTGAGCCACATCCGTTTGGATTCCTGTCAGATGCAACAGATTTG 360
DB 617 AATATATACCTGAGCCACATCCGTTTGGATTCCTGTCAGATGCAACAGATTTG 676
OY 361 ATGAGGTCGAGGTTTACACTGAGAGCCCACTTATCATGATTTATGGAAGAAATGAAG 420
DB 677 ATGAGGTCGAGGTTTACACTGAGAGCCCACTTATCATGATTTATGGAAGAAATGAAG 736
OY 421 CTAGTAGACTCCAGAGATGTCAGCTCATCAAGATTTTACAATACTTACAGATCCACCG 480
DB 737 CTAGTAGACTCCAGAGATGTCAGCTCATCAAGATTTTACAATACTTACAGATCCACCG 796
OY 481 GGGCCCCAGGGGTCACAGAGGTGACAGAGATCCAGAGACCCCTGAGCCCAACTGCAAC 540
DB 797 GGGCCCCAGGGGTCACAGAGGTGACAGAGATCCAGAGACCCCTGAGCCCAACTGCAAC 856
OY 541 AAGGACAGAAAGAGAGAGAGAGGAGCGCTGAGACACCTGCGCTCGGGGTAGAGAGGC 600
DB 857 AAGGACAGAAAGAGAGAGAGAGGAGCGCTGAGACACCTGCGCTCGGGGTAGAGAGGC 916
OY 601 CCAATTTGAGACCAAGCTGTGCTCCCGGAGAGCGTGGGCGGCAAAAGATCTAAAGCTCCAG 660
DB 917 CCAATTTGAGACCAAGCTGTGCTCCCGGAGAGCGTGGGCGGCAAAAGATCTAAAGCTCCAG 976
OY 661 GGGCCCCAAGAGCTCCCGTGTGCTCCCGGAGAGCGTGGGCGGCAAAAGATCTAAAGCTCCAG 720
DB 977 GGGCCCCAAGAGCTCCCGTGTGCTCCCGGAGAGCGTGGGCGGCAAAAGATCTAAAGCTCCAG 1036
OY 721 CCAAGGCCCCCGGGCCACACAGGCAAAAGAGAGCGTCCCGGCGCTCAGAGGCCCTCTGAGC 780
DB 1037 CCAAGGCCCCCGGGCCACACAGGCAAAAGAGAGCGTCCCGGCGCTCAGAGGCCCTCTGAGC 1096
OY 781 TTCCAGAGGACTTCAGAGGACCGTTGGGAGAGCGTGGGCGGCTCTGAGCTCGGGAGCTGCCA 840
DB 1097 TTCCAGAGGACTTCAGAGGACCGTTGGGAGAGCGTGGGCGGCTCTGAGCTCGGGAGCTGCCA 1156
OY 841 GGGCTTGGCTGGGGTACAGAGCATGCCAGGCCCCCAAGGGCCCCCGGCTCTGAGGCCA 900
DB 1157 GGGCTTGGCTGGGGTACAGAGCATGCCAGGCCCCCAAGGGCCCCCGGCTCTGAGGCCA 1216
OY 901 TCAGAGAGCGGTGTGCTGCTGCGCTGCGCTGAGAGATGAGACCAACCCCGGACCGAGAGCAAT 960
DB 1217 TCAGAGAGCGGTGTGCTGCTGCGCTGCGCTGAGAGATGAGACCAACCCCGGACCGAGAGCAAT 1276
OY 961 GGGTGGCCCCCTCAGCTGGAAGAACTTCACAGCAAAATGCTACTATTTTTCAGTTGGA 1020
DB 1277 GGGTGGCCCCCTCAGCTGGAAGAACTTCACAGCAAAATGCTACTATTTTTCAGTTGGA 1336
OY 1021 GAAATTTTTCAGAGATGCAAAAGCTTTTCTGTGAAGACAGCTTCACATCTTGTTCAT 1080
DB 1337 GAAATTTTTCAGAGATGCAAAAGCTTTTCTGTGAAGACAGCTTCACATCTTGTTCAT 1396
OY 1081 AACACTAGAGAGAGAGAGCAATGAGTAAAAAAGATGCTAGGAGAGAGAGAGCACTGG 1140
DB 1397 AACACTAGAGAGAGAGAGCAATGAGTAAAAAAGATGCTAGGAGAGAGAGAGCACTGG 1456
OY 1141 ATCGGCTCAGAGACTCAGAGGCTGAATAATGATGAATGAGTGGCTGATGGAGCATCTCCA 1200

Db 1457 ATGGGCTCAGACACTGAGCGTGAATAAGAAAGCGCTGGATGGACATCTCCA 1516
 QY 1201 GACTACAAAATTTGGAAAGTGGACACCGGATTAAGTGGGTCATGGCCATGGCCCGCA 1260
 Db 1517 GACTACAAAATTTGGAAAGTGGACACCGGATTAAGTGGGTCATGGCCATGGCCCGCA 1576
 QY 1261 GAAGACTGTGCTGGTGTGATTTATGCTGGCAGTGTGAACGATTTCCAAATGTGAAGACGTC 1320
 Db 1577 GAAGACTGTGCTGGTGTGATTTATGCTGGCAGTGTGAACGATTTCCAAATGTGAAGACGTC 1636
 QY 1321 AATTAATCTTATTTGCGAAAAAGACAGGAGACAGTACTGATCTGCATTA 1371
 Db 1637 AATTAATCTTATTTGCGAAAAAGACAGGAGACAGTACTGATCTGCATTA 1687
 RESULT 5
 AAC66903
 ID AAC66903 standard; cDNA; 2641 BP.
 AC AAC66903;
 XX
 DT 27-MAR-2001 (first entry)
 DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
 XX
 XX Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 KM inflammation; reproductive disorder; cardiovascular disorder;
 KM immune disorder; musculoskeletal disorder; developmental disorder;
 KM gastrointestinal disorder; cell proliferation disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200068380-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000MO-US12811.
 XX
 PR 11-MAY-1999; 99US-0133643.
 PR 23-AUG-1999; 99US-0150409.
 XX
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Bandman O, Hillman JL, Tang YF, Lai P, Yue H, Baughn MR, Lu DM;
 PI Azimzai Y;
 XX
 DR WPI: 2001-007395/01.
 DR P-PSDB; AAB27236.
 PT Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX
 PS Claim 4; Page 121-122; 129pp; English.
 XX
 XX The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 XX
 SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;
 Query Match 99.9%; Score 1369.4; DB 22; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAACAGCCAGCTCAACTCATTACAGAGTCCAGATGGAGAAATCAACACTATCTCTCAA 60
 Db 567 ATGAACAGCCAGCTCAACTCATTACAGAGTCCAGATGGAGAAATCAACACTATCTCTCAA 626
 QY 61 GCCAAGACAGAGAAACCTGGAAGACCTGCAAGACTTACACAAATATGAGAGATTAACAA 120
 Db 627 GCCAAGACAGAGAAACCTGGAAGACCTGCAAGACTTACACAAATATGAGAGATTAACAA 686
 QY 121 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTCTTGTAGAGCGATATTGTGAAC 180
 Db 687 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTCTTGTAGAGCGATATTGTGAAC 746
 QY 181 ATCATTAAGCAATATCAATTACACAGCCACACCTCGGAGCGTGAACAGCAATCTAAT 240
 Db 747 ATCATTAAGCAATATCAATTACACAGCCACACCTCGGAGCGTGAACAGCAATCTAAT 806
 QY 241 GAAGTCAAGACCACTTGCACAGATACCCCTTACCAAAACACAGATATGACCTCTG 300
 Db 807 GAAGTCAAGACCACTTGCACAGATACCCCTTACCAAAACACAGATATGACCTCTG 866
 QY 301 AATAATACCTGAGCAACATCCGTTGGATTTCTCTCAGATGCAACAGATTTG 360
 Db 867 AATAATACCTGAGCAACATCCGTTGGATTTCTCTCAGATGCAACAGATTTG 926
 QY 361 ATGAGGTGAGGTTAGACACTGAAGTGAAGCCACTTATCACTGATATGAAGAATGAAG 420
 Db 927 ATGAGGTGAGGTTAGACACTGAAGTGAAGCCACTTATCACTGATATGAAGAATGAAG 986
 QY 421 CTGATGACCTCCAAAGCATGCTGATCAATCAAAATTTTCAATACAGAGTCCACCG 480
 Db 987 CTGATGACCTCCAAAGCATGCTGATCAATCAAAATTTTCAATACAGAGTCCACCG 1046
 QY 481 GGCCCCAGGGGTCCAAAGAGTGAAGAGATGCCAGAGACCCCTGAGCCCAACTGGCAAC 540
 Db 1047 GGCCCCAGGGGTCCAAAGAGTGAAGAGATGCCAGAGACCCCTGAGCCCAACTGGCAAC 1106
 QY 541 AAGGACAGAAAGAGAGAAAGGGGAGCTGACACACTGCGCTCGGGTGAAGAGC 600
 Db 1107 AAGGACAGAAAGAGAGAAAGGGGAGCTGACACACTGCGCTCGGGTGAAGAGC 1166
 QY 601 CCAATTGACACAGCTGGTCCCGGAGAGCTGGGGGCAAGAGATCTAAGGCTCCGAC 660
 Db 1167 CCAATTGACACAGCTGGTCCCGGAGAGCTGGGGGCAAGAGATCTAAGGCTCCGAC 1226
 QY 661 GGCCCCAAAGGCTCCGCTGGTCCCTGGGAGACCCGCTCAAGGACCCCACTGGGAGC 720
 Db 1227 GGCCCCAAAGGCTCCGCTGGTCCCTGGGAGACCCGCTCAAGGACCCCACTGGGAGC 1286
 QY 721 CCAAGGCCCCGGGGCCACAGGCAAAAGAGACTCCCGGCTCAAGGACCTCTGAGC 780
 Db 1287 CCAAGGCCCCGGGGCCACAGGCAAAAGAGACTCCCGGCTCAAGGACCTCTGAGC 1346
 QY 781 TTCAGAGGACTTCAAGGACACGTTGGGAGCTGCGGGTCTGAGCTCGAGGAGTGC 840
 Db 1347 TTCAGAGGACTTCAAGGACACGTTGGGAGCTGCGGGTCTGAGCTCGAGGAGTGC 1406
 QY 841 GGCTTGGCTGGGTATACAGGATGCAAGGCCCAAGGGCCCCCGGCTCTGAGCCCA 900
 Db 1407 GGCTTGGCTGGGTATACAGGATGCAAGGCCCAAGGGCCCCCGGCTCTGAGCCCA 1466
 QY 901 TCAGAGAGCGGTGCTGCCCCCTGCGCTGCAAGATGAGACCAACCCGGGACCGGAGACAT 960
 Db 1467 TCAGAGAGCGGTGCTGCCCCCTGCGCTGCAAGATGAGACCAACCCGGGACCGGAGACAT 1526
 QY 961 GGCTGGCCCGGCTCACTGGAAGAACTTCAAGCAAAATGCTACTATTTTCAAGTTGAGAA 1020
 Db 1527 AGCTGCCCGGCTCACTGGAAGAACTTCAAGCAAAATGCTACTATTTTCAAGTTGAGAA 1586
 QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGGAAGACAGCTTCACATCTTGTTCATA 1080
 Db 1587 GAAATTTTGGAGATGCAAAAGCTTTTCTGGAAGACAGCTTCACATCTTGTTCATA 1646
 QY 1081 AACACTAGAGAGAAACAGCAATGATTAATAAAACAGATGTTAGGAGAGAGACCTGG 1140

QY	541	AAGGACAGAAAGAGAGAAAGGGGAGAGCCTGGACCACTGGCCCTGGGGGTGAGAGAGGC	600
Dp	1461	AAGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACTGGCCCTGGGGGTGAGAGAGGC	1520
QY	601	CCAAATTGGACACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAAAGATCTTAAAGGGCTCCAG	660
Dp	1521	CCAAATTGGACACAGCTGGTCCCCCGGAGAGCGTGGGGCAAAAGATCTTAAAGGGCTCCAG	1580
QY	661	GGCCCCAAAGGCTTCCTGGTTCCTCTGGGAAGCCGGCCTCAGGGCCCCAGTGGGAC	720
Dp	1581	GGCCCCAAAGGCTTCCTGGTTCCTCTGGGAAGCCGGCCTCAGGGCCCCAGTGGGAC	1640
QY	721	CCAGGGCCCCCGGGCCACCACAGGCAAAAGAGGAGACTCCCGGGCCTCAGAGGCCCTCTGGAC	780
Dp	1641	CCAGGGCCCCCGGGCCACCACAGGCAAAAGAGGAGACTCCCGGGCCTCAGAGGCCCTCTGGAC	1700
QY	781	TTCCAGGAGACTTCAGGGACACCCTTGGGGAGCCTGGGGTGCCTGAGACCTGGGGAGCTGGCA	840
Dp	1701	TTCCAGGAGACTTCAGGGACACCCTTGGGGAGCCTGGGGTGCCTGAGACCTGGGGAGCTGGCA	1760
QY	841	GGCTTGCTGGGGGATCCAGGCAATGGCCACAGGCCCCCGGGCCCTCGTGGGCCA	900
Dp	1761	GGCTTGCTGGGGGATCCAGGCAATGGCCACAGGCCCCCGGGCCCTCGTGGGCCA	1820
QY	901	TCAGAGACGGTGGTGCCCTGGGCTGTCAGAAATGAGACCAACCCGGCACCGAGAGACAAT	960
Dp	1821	TCAGAGACGGTGGTGCCCTGGGCTGTCAGAAATGAGACCAACCCGGCACCGAGAGACAAT	1880
QY	961	GGCTGCCCGCTCACTGGAGAACTTTCACAGACCAATGCTACTATTTTTCAGTGGAGAA	1020
Dp	1881	AGCTGCCCGCTCACTGGAGAACTTTCACAGACCAATGCTACTATTTTTCAGTGGAGAA	1940
QY	1021	GAATTTTGTAGAGTGTGAAGCTTTCTGTGAAGCAACTCTTCACATCTGTGTTCATA	1080
Dp	1941	GAATTTTGTAGAGTGTGAAGCTTTCTGTGAAGCAACTCTTCACATCTGTGTTCATA	2000
QY	1081	AACACTGAGAGGAGAACGCAATGGATMAAAACACAGATGGTGGGAGAGAGGCACTGG	1140
Dp	2001	AACACTGAGAGGAGAACGCAATGGATMAAAACACAGATGGTGGGAGAGAGGCACTGG	2060
QY	1141	ATCGGGCTCACACACTCAGACGCTGAAATTAATGATGAGTAAGTGGCTGGATGGACATCTCCA	1200
Dp	2061	ATCGGGCTCACACACTCAGACGCTGAAATTAATGATGAGTAAGTGGCTGGATGGGACATCTCCA	2120
QY	1201	GACTACAAAAAATTGGAAAGCTGTGACAGCCGGATTAATCTGGGGTCATGGGCAATGGGCCACAGA	1260
Dp	2121	GACTACAAAAAATTGGAAAGCTGTGACAGCCGGATTAATCTGGGGTCATGGGCAATGGGCCACAGA	2180
QY	1261	GAAAGCTGTGCTGGGTTGATTATTCGTGGGCACTGGAAACGATTTTCCAATGTGAAGAGCTC	1320
Dp	2181	GAAAGCTGTGCTGGGTTGATTATTCGTGGGCACTGGAAACGATTTTCCAATGTGAAGAGCTC	2240
QY	1321	AATTAACCTCAATTTTGGCAAAAAGCAGGAGAGACAGTACGTCTATCTTGCATTA	1371
Dp	2241	AATTAACCTCAATTTTGGCAAAAAGCAGGAGAGACAGTACGTCTATCTTGCATTA	2291

Accession	Gene	Protein	Species	Source	Reference
AAV55746	23-MAR-1999 (first entry)	Human secreted protein clone bv227_1 coding sequence.	Human	Secreted protein; human; nutritional supplements; immune stimulant; immune suppressor; haemotopiasis regulator; tissue growth; chemotaxis activin/inhibin; chemokinesis; haemostasis; thrombolysis; receptor/ligand activity; anti-inflammatory; tumour inhibitor; cadherin/tumour invasion suppressor; ds.	

XX	Homo sapiens.
OS	
XX	
FH	Key Location/Qualifiers
FT	67..693
ET	/tag= a
XX	
PN	W09855614-A2.
XX	
PD	10-DEC-1998.
XX	
PF	01-JUN-1998; 98WO-US11210.
XX	
PR	29-MAY-1998; 98US-0087255.
XX	
PR	04-JUN-1997; 97US-0868696.
PR	04-JUN-1997; 97US-0868697.
PR	04-JUN-1997; 97US-0868698.
PR	04-JUN-1997; 97US-0868698.
PR	04-JUN-1997; 97US-0868699.
PR	04-JUN-1997; 97US-0868900.
PR	04-JUN-1997; 97US-0869191.
PR	04-JUN-1997; 97US-0869192.
PR	04-JUN-1997; 97US-0869193.
PR	04-JUN-1997; 97US-0869194.
PA	(GENY) GENETICS INST INC.
XX	
PI	Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
PI	Mccoy JM, Racie LA, Spaulding V, Treacy M;
XX	
DR	WP1. 1999-059912/05.
XX	P-P5DB; AAM73628.
XX	
PT	New polynucleotides encoding secreted human proteins - derived from
PT	human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT	retina, adult placenta or adult uterus CDNA libraries
XX	
PS	Claim 26; Page 87-88; 127pp; English.
CC	
CC	This sequence encodes a human secreted protein of the invention.
CC	This DNA sequence was isolated from a human adult brain cDNA
CC	library, and was designated clone bv227.1. The DNAs and proteins
CC	are predicted to have biological activities which would make them
CC	suitable for treating, preventing or ameliorating medical conditions in
CC	humans and animals, although no supporting data is given. Suggested
CC	activities include nutritional sources or supplements, immune
CC	stimulating or suppressing activity, haematopoiesis regulating activity,
CC	tissue growth activity, activin/inhibin activity,
CC	cytoretic/chemokine activity, haemostatic and thrombolytic activity,
CC	receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC	invasion suppressor activity, and tumour inhibition activity. The DNAs
CC	are also stated to be useful for gene therapy. A host cell transfected
CC	with the DNA or its subsegments and variants is useful for recombinant
CC	production of the human secreted protein clones.
XX	
SO	Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other:

Query Match	Similarity	99.9%	Score	1369.4	DB	20	Length	2930	
Best Local	Similarity	99.9%	Pred. No.	0					
Matches	1370	Conservative	0	Mismatches	1	Indels	0	Gaps	0
QY	1	ATGAACAGCAGCCTCAACTCATTTCACAGGTCCAGATGAGAGAACATCAGCCACTATCTCTCAA							60
Db	921	ATGAACAGCAGCCTCAACTCATTTCACAGGTCCAGATGAGAGAACATCAGCCACTATCTCTCAA							980
QY	61	GCCAAACGAGCAGACCTTGAAAGACCTGCGAGGACTTACACAAAGATGCAGAGATATGAGACA							120
Db	981	GCCAAACGAGCAGACCTTGAAAGACCTGCGAGGACTTACACAAAGATGCAGAGATATGAGACA							1040
QY	121	GCCATCAAGTTCAACCAACTGGAGAGACGCTCCAGCTCTTTGAGACGGATATTGTGAAAC							180
Db	1041	GCCATCAAGTTCAACCAACTGGAGAGACGCTCTCCAGCTCTTTGAGACGGATATTGTGAAAC							1100

QY 181 ATCATTTAGCAATATTCAGTTTACACAGCCACACCTGCGGAGCGCTGACACCAATCTAAAT 240
 Db 1101 ATCATTTAGCAATATTCAGTTTACACAGCCACACCTGCGGAGCGCTGACACCAATCTAAAT 1160
 QY 241 GAAGTAGAGACCACTTGACACAGTACCCTTACAAAACACACAGATGATCTGACCTCTTG 300
 Db 1161 GAAGTAGAGACCACTTGACACAGTACCCTTACAAAACACACAGATGATCTGACCTCTTG 1220
 QY 301 AATTAATACCTGCGCCAAACATCCGTTTGATCTGTTCTCTGAGATGCAACAAGATTG 360
 Db 1221 AATTAATACCTGCGCCAAACATCCGTTTGATCTGTTCTCTGAGATGCAACAAGATTG 1280
 QY 361 ATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGTATGGAAGAAATGAAG 420
 Db 1281 ATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGTATGGAAGAAATGAAG 1340
 QY 421 CTAGTAGACTCCACAGCATGCTGCTCATCAGAAATTTTACAACTACTACAGGTCACCG 480
 Db 1341 CTAGTAGACTCCACAGCATGCTGCTCATCAGAAATTTTACAACTACTACAGGTCACCG 1400
 QY 481 GGGCCCCAGGGGTCACAGAGGTGACAGAGATCCAGGAGCCCTGCGCCCACTGGCAAC 540
 Db 1401 GGGCCCCAGGGGTCACAGAGGTGACAGAGATCCAGGAGCCCTGCGCCCACTGGCAAC 1460
 QY 541 AAGGACAGAAAAGAGAGAGAGGAGGAGCCTGACACACTGCGCCCTGCGGTCAGAGAGGC 600
 Db 1461 AAGGACAGAAAAGAGAGAGAGGAGGAGCCTGACACACTGCGCCCTGCGGTCAGAGAGGC 1520
 QY 601 CCAATTTGACACAGCTGCTCCCCCGAGAGACCTGCGGCAAGAGATCTAAAGGCTCCAG 660
 Db 1521 CCAATTTGACACAGCTGCTCCCCCGAGAGACCTGCGGCAAGAGATCTAAAGGCTCCAG 1580
 QY 661 GGGCCCCAAGGGTCCCGTGGTTCCTGGAAGCCCGCCCTCAGAGGCCCCCACTGGGAGC 720
 Db 1581 GGGCCCCAAGGGTCCCGTGGTTCCTGGAAGCCCGCCCTCAGAGGCCCCCACTGGGAGC 1640
 QY 721 CCAGGCCCCCGGAGCCACACAGGCAAAAGAGAGGAGCTCCCGGCTCAGAGGCCCCCTGCGC 780
 Db 1641 CCAGGCCCCCGGAGCCACACAGGCAAAAGAGAGGAGCTCCCGGCTCAGAGGCCCCCTGCGC 1700
 QY 781 TTCCAGGAGCTTCAGGAGGACCCGTTGGGAGCCCTGAGGAGCTGCGGAGCTGCCA 840
 Db 1701 TTCCAGGAGCTTCAGGAGGACCCGTTGGGAGCCCTGAGGAGCTGCGGAGCTGCCA 1760
 QY 841 GGGTTCCTGGGGTTACAGGAGTCCAGGCCCCCAAGGGCCCCCGGCTCTGCGGCCA 900
 Db 1761 GGGTTCCTGGGGTTACAGGAGTCCAGGCCCCCAAGGGCCCCCGGCTCTGCGGCCA 1820
 QY 901 TCAGAGAGCGGTGTCGCCCTGCGCTGAGCAATGAGCAACCCCGGACCGAGAGCAAT 960
 Db 1821 TCAGAGAGCGGTGTCGCCCTGCGCTGAGCAATGAGCAACCCCGGACCGAGAGCAAT 1880
 QY 961 GGGTCCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
 Db 1881 GGGTCCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1940
 QY 1021 GAAATTTTGGAGGAGCAAGGCTTCTGTAAGAGACAACTTCACATCTGTTTTCATA 1080
 Db 1941 GAAATTTTGGAGGAGCAAGGCTTCTGTAAGAGACAACTTCACATCTGTTTTCATA 2000
 QY 1081 AACACTAGAGAGAGAACAGCAATGATTAATAAAACAGATGATGAGGAGAGAGACCACTGG 1140
 Db 2001 AACACTAGAGAGAGAACAGCAATGATTAATAAAACAGATGATGAGGAGAGAGACCACTGG 2060
 QY 1141 ATCGGCTCAGACACTCAAGCGTGAATAATGAATGGAAGTGGCTGATGGAGCATCTCCA 1200
 Db 2061 ATCGGCTCAGACACTCAAGCGTGAATAATGAATGGAAGTGGCTGATGGAGCATCTCCA 2120
 QY 1201 GACTACAAAATTTGAAAACCTGGACAGCGGAGTAACTGGGGTATGAGGAGGAGGAGG 1260
 Db 2121 GACTACAAAATTTGAAAACCTGGACAGCGGAGTAACTGGGGTATGAGGAGGAGGAGG 2180
 QY 1261 GAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCATTTCCAAATGTGAAGAGCTC 1320

Db 2181 GAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAACGATTTCCAAATGTGAAGAGCTC 2240
 QY 1321 AATACTTCAATTTGCGAAAAAGACAGGAGACAGTACTGTCATCTGATTA 1371
 Db 2241 AATACTTCAATTTGCGAAAAAGACAGGAGACAGTACTGTCATCTGATTA 2291
 RESULT 8
 ID AA158842 standard; cDNA; 2318 BP.
 AC
 XX
 AA158842;
 XX
 DT 22-OCT-2001 (first entry) -
 XX
 DE Human polynucleotide SEQ ID NO 1045.
 XX
 KW Human; nootropic; immunosuppressant; cytosolic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 PD
 XX 26-JUL-2001.
 PF
 XX 26-DEC-2000; 2000MO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Dirmanac RJ;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM39686.
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1045; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,
 CC immunosuppressant and cytosolic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Actin/inhibn activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Query Match 99.1%; Score 1358.4; DB 22; Length 2318;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1370; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGACAGCCAGCTCACTATTACAGTACAGTGAAGACATCACCATTCTTCTCAA 60
DB 317 ATGACAGCCAGCTCACTATTACAGTACAGTGAAGACATCACCATTCTTCTCAA 376
QY 61 -GCCAAGCAGACAGTCAAGTGAAGAGCTGACAGTCTACACAAAGATGACAGATAGAAC 119
DB 377 GGCCAGACAGACAGTCAAGTGAAGAGCTGACAGTCTACACAAAGATGACAGATAGAAC 436
QY 120 AGCCATCAAGTTCACCACTGAGAACCTCTTCAGCTCTTTGAGACGATATTGTGAA 179
DB 437 AGCCATCAAGTTCACCACTGAGAACCTCTTCAGCTCTTTGAGACGATATTGTGAA 496
QY 180 CATCATTAAGCAATATCAGTATACAGACCCACACCTGCGAGAGCTGACAGCATCTTAA 239
DB 497 CATCATTAAGCAATATCAGTATACAGACCCACACCTGCGAGAGCTGACAGCATCTTAA 556
QY 240 TGAAGTACAGACACCTGACAGATACCTTACCAACACACAGATATGTCGACTCTT 299
DB 557 TGAAGTACAGACACCTGACAGATACCTTACCAACACACAGATATGTCGACTCTT 616
QY 300 GAATTAATACCTGCGCAACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATT 359
DB 617 GAATTAATACCTGCGCAACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATT 676
QY 360 GATGAGTCCAGGTAAACACTGAGTAGCCAACTTATCAGTGTATTTGAGAAATGAA 419
DB 677 GATGAGTCCAGGTAAACACTGAGTAGCCAACTTATCAGTGTATTTGAGAAATGAA 736
QY 420 GCTAGTACTCCAGAGATGTCAGTCAATCAAGATTTTACATPACTACAAAGTCCACC 479
DB 737 GCTAGTACTCCAGAGATGTCAGTCAATCAAGATTTTACATPACTACAAAGTCCACC 796
QY 480 GGGCCCCAGGGGTCCAAAGAGTGACAGAGATCCCAAGGACCCCTGGCCCCACTGSCAA 539
DB 797 GGGCCCCAGGGGTCCAAAGAGTGACAGAGATCCCAAGGACCCCTGGCCCCACTGSCAA 856
QY 540 CAAGGGACAAAGAGAGAGAGGGGAGCCTGGACCACTGGCCCTCGGGTGAAGAG 599
DB 857 CAAGGGACAAAGAGAGAGAGGGGAGCCTGGACCACTGGCCCTCGGGTGAAGAG 916
QY 600 CCCAATTGGACACAGCTGTCCCGCCGAGAGAGCTGGCGGCAAAAGATCTAAAGCTCCCA 659
DB 917 CCCAATTGGACACAGCTGTCCCGCCGAGAGAGCTGGCGGCAAAAGATCTAAAGCTCCCA 976
QY 660 GGGCCCCAAAGGCTCCGCTGGTTCCCTGGGAAAGCCCGGCCCTCAGAGGCCCAAGTGGGA 719
DB 977 GGGCCCCAAAGGCTCCGCTGGTTCCCTGGGAAAGCCCGGCCCTCAGAGGCCCAAGTGGGA 1036
QY 720 CCCAGGCCCCCGGGCCCCACCAAGGAAAGAGGACTCCCGGGCCCTCAGAGGCCCTCTGG 779
DB 1037 CCCAGGCCCCCGGGCCCCACCAAGGAAAGAGGACTCCCGGGCCCTCAGAGGCCCTCTGG 1096
QY 780 CTTCCAGAGGACTTCAGGGACACCGTTGGGAGAGCTGGGGTGCCTGAGACTCGGGGAGCTGCC 839
DB 1097 CTTCCAGAGGACTTCAGGGACACCGTTGGGAGAGCTGGGGTGCCTGAGACTCGGGGAGCTGCC 1156
QY 840 AGGCTTCCTGGGGTACCAAGGATCCAGGACCCCAAGGGGCCCCCGGCCCTCTCTGGCCC 899
DB 1157 AGGCTTCCTGGGGTACCAAGGATCCAGGACCCCAAGGGGCCCCCGGCCCTCTCTGGCCC 1216
QY 900 ATCAGAGAGCGGTGGTCCCTGGCCCTGCAAGATGAGCCCAACCCGGCAGCGAGAGCAA 959
DB 1217 ATCAGAGAGCGGTGGTCCCTGGCCCTGCAAGATGAGCCCAACCCGGCAGCGAGAGCAA 1276
QY 960 TGGCTGCCCGCTCACTGGAAGAACTCACAGCAAAATGCTACTATTCTTTCAGTTGAGAA 1019
DB 1277 TGGCTGCCCGCTCACTGGAAGAACTCACAGCAAAATGCTACTATTCTTTCAGTTGAGAA 1336

QY 1020 AGAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAGTCTTCACATCTTTTTCAT 1079
DB 1337 AGAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAGTCTTCACATCTTTTTCAT 1396
QY 1080 AAACACTAGAGAGACAGCAATGGATAAAAAACACATGTTAGGAGAGAGAGCCACTG 1139
DB 1397 AAACACTAGAGAGAGACAGCAATGGATAAAAAACACATGTTAGGAGAGAGAGCCACTG 1456
QY 1140 GATGCGCTCAGACACTCAGAGCTGAAAGATGAATGAAGTGGCTGATGGAGCATCTCC 1199
DB 1457 GATGCGCTCAGACACTCAGAGCTGAAAGATGAATGAAGTGGCTGATGGAGCATCTCC 1516
QY 1200 AGACTACAAAATTTGAAAGCTGACAGCCGGATTAAGTGGGGTCAATGGCCAGG 1259
DB 1517 AGACTACAAAATTTGAAAGCTGACAGCCGGATTAAGTGGGGTCAATGGCCAGG 1576
QY 1260 AGAAGCTGCTGGGTGATTTATGCTGGGCACTGGAACGATTTCCAAATGTGAAGCT 1319
DB 1577 AGAAGCTGCTGGGTGATTTATGCTGGGCACTGGAACGATTTCCAAATGTGAAGCT 1636
QY 1320 CAATTAATCTATTTGCGAAAAAGACAGAGAGACAGTACTGTATCTCATTTA 1371
DB 1637 CAATTAATCTATTTGCGAAAAAGACAGAGAGACAGTACTGTATCTCATTTA 1688

RESULT 9
ABL95574
ID ABL95574 standard; cDNA; 2005 BP.
XX
AC ABL95574;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.
XX
KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiant; cytoslastic; antiangiogenic; hypotensive; vulnerary;
KW antiarteriosclerotic; gene; ss.
XX
OS Homo sapiens.
PN WO200208284-A2.
XX
PD 31-JAN-2002.
XX
PF 09-JUL-2001; 2001MO-US21735.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000MO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000MO-US23522.
PR 24-AUG-2000; 2000MO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000MO-US30952.
PR 10-NOV-2000; 2000MO-US30873.
PR 01-DEC-2000; 2000MO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000MO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001MO-US06520.
PR 01-MAR-2001; 2001MO-US06666.
PR 09-MAR-2001; 2001US-0802706.

14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001US-0866034.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001US-0870574.
PR 01-JUN-2001; 2001US-0870574.
PR 20-JUN-2001; 2001US-0870574.
PR 28-JUN-2001; 2001US-0870574.
XX
PA (GENE) GENENTECH INC.
PA (BAKE) BAKER K P.
PA (FERB) FERRARA N.
PA (GERB) GERBER H.
PA (GERR) GERRITSEN M E.
PA (GODD) GODDARD A.
PA (GODO) GODOWSKI P J.
PA (GURN) GURNEY A L.
PA (HILL) HILLMAN K J.
PA (MARS) MARSTERS S A.
PA (PANJ) PAN J.
PA (PAON) PAONI N F.
PA (STEP) STEPHAN J F.
PA (WATA) WATANABE C K.
PA (WILL) WILLIAMS P M.
PA (WOOD) WOOD W I.
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillman KJ, Marsters SA, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
DR WPI: 2002-171999/22.
DR P-PSDB: ABB95436.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
XX
PS Claim 1; Fig 27; 567pp; English.
XX
XX The present invention provides the protein and coding sequences of human
XX PRO proteins. These are useful for treating or diagnosing a
XX cardiovascular, endothelial or angiogenic disorder, including cardiac
XX hypertrophy, trauma, cancer, age-related macular degeneration,
XX atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
XX angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
XX angiogenesis (such as breast carcinoma and liver carcinoma) and wound
XX healing. The present sequence is a coding sequence of the invention.
XX
XX Sequence 2005 BP: 606 A; 493 C; 491 G; 415 T; 0 other;

Query Match 99.0%; Score 1357.4; DB 24; Length 2005;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGAACAGCCAGCTCACTCATTCACAGGTGAGTGAAGACATCACTATCTCTCA 60
DB 134 ATGAACAGCCAGCTCACTCATTCACAGGTGAGTGAAGACATCACTATCTCTCA 193
QY 61 GCCAAGGAGCGAAGCTGAAGACTGCGAGCTTACACAAAGATGACAGATGAGACA 120
DB 134 GCCAAGGAGCGAAGCTGAAGACTGCGAGCTTACACAAAGATGAGATGAGACA 253
QY 121 GCCATCAAGTTCAACCACTGAGAGACGCTCCAGCTCTTGAGACGGATTTTGAAAC 180
DB 254 GCCATCAAGTTCAACCACTGAGAGACGCTCCAGCTCTTGAGACGGATTTTGAAAC 313
QY 161 ATCATTAACAATATGATTACAGCCACCAACCTGCGAGCGCTGACCAATCTAAT 240

DB 314 ATCATTAACAATATGATTACAGCCACCACTGCGAGCGCTGACCAATCTAAT 373
QY 241 GAAGTCAGAGACCACTTGCACATACCCCTTACCAACACAGATGATGACCTGCTG 300
DB 374 GAAGTCAGAGACCACTTGCACATACCCCTTACCAACACAGATGATGACCTGCTG 432
QY 301 AATTAATACCTTGGCCCAACATCCGTTTGGATTTGTTTCTCAGATGCAACAAGATTG 360
DB 433 AATTAATACCTTGGCCCAACATCCGTTTGGATTTGTTTCTCAGATGCAACAAGATTG 492
QY 361 ATGAGTCGAGGTTGAGACACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420
DB 493 ATGAGTCGAGGTTGAGACACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 552
QY 421 CTAGTAGACTCCAAAGCATGTGTAGCTCATCAAGATTTTACATTAAGTTCACCG 480
DB 553 CTAGTAGACTCCAAAGCATGTGTAGCTCATCAAGATTTTACATTAAGTTCACCG 612
QY 481 GGCCCCAGGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCTGCGCCACTGGCAAC 540
DB 613 GGCCCCAGGGGTCCCAAGAGGTGACAGAGATCCCAAGGACCCCTGCGCCACTGGCAAC 672
QY 541 AAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 673 AAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 732
QY 601 CCAATTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 733 CCAATTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 792
QY 661 GGCCCCAAAGAGCTCCGTTGTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 793 GGCCCCAAAGAGCTCCGTTGTTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 852
QY 721 CCAGGCCCCCGGGCCACAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 853 CCAGGCCCCCGGGCCACAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 912
QY 781 TTCCAGGAGACTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 913 TTCCAGGAGACTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972
QY 841 GGCTTGCTGCGGAGTACAGAGCATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 973 GGCTTGCTGCGGAGTACAGAGCATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1032
QY 901 TCAGAGAGCGGTGTCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 960
DB 1033 TCAGAGAGCGGTGTCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1092
QY 961 GGCTGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1020
DB 1093 GGCTGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1152
QY 1021 GAAATTTTGGAGATGCAAGCTTTTCTGTAAGACAAAGTCTTCAATCTTTTTCATA 1080
DB 1153 GAAATTTTGGAGATGCAAGCTTTTCTGTAAGACAAAGTCTTCAATCTTTTTCATA 1212
QY 1081 AACACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1213 AACACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1272
QY 1141 ATCGGCTTCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1273 ATCGGCTTCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1332
QY 1201 GACTACAAAATTTGAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1333 GACTACAAAATTTGAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1392
QY 1261 GAAGACTGTGTGGGTGATTTATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1393 GAAGACTGTGTGGGTGATTTATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1452

QY 1321 AATACTGATTTTGCAGAAAAGACAGGACAGACTGTCATCTGCATTA 1371
DB 1453 AATAACTGATTTGCGAAAAAGACAGGAGACGACTGTCATCTGCATTA 1503
RESULT 10
ABL88085
ID ABL88085 standard; cDNA; 2005 BP.
XX
AC ABL88085;
XX
DT 16-MAY-2002 (first entry)
XX
DE Human PRO7223 cDNA sequence SEQ ID NO:27.
XX
KW Human; angiogenesis; cardiact; cyostatic; antilangogenic; hypotensive;
KW vulnerary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW anglogenetic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial restenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200690-A2.
XX
PD 03-JAN-2002.
XX
PF 20-JUN-2001; 2001WO-US19692.
XX
XX 23-JUN-2000; 2000US-213637P.
XX 20-JUL-2000; 2000US-219556P.
XX 25-JUL-2000; 2000US-220624P.
XX 25-JUL-2000; 2000US-220644P.
XX 28-JUL-2000; 2000WO-US20710.
XX 02-AUG-2000; 2000US-222695P.
XX 17-AUG-2000; 2000US-0643657.
XX 23-AUG-2000; 2000WO-US23522.
XX 24-AUG-2000; 2000WO-US23328.
XX 07-SEP-2000; 2000US-230978P.
XX 18-SEP-2000; 2000US-0664610.
XX 18-SEP-2000; 2000US-0665350.
XX 18-SEP-2000; 2000US-242922P.
XX 24-OCT-2000; 2000US-0709238.
XX 08-NOV-2000; 2000US-0709238.
XX 08-NOV-2000; 2000WO-US30952.
XX 10-NOV-2000; 2000WO-US30873.
XX 01-DEC-2000; 2000WO-US32878.
XX 20-DEC-2000; 2000US-0747259.
XX 20-DEC-2000; 2000WO-US34956.
XX 22-JAN-2001; 2001US-0767609.
XX 28-FEB-2001; 2001US-0796498.
XX 28-FEB-2001; 2001WO-US06520.
XX 01-MAR-2001; 2001WO-US06666.
XX 09-MAR-2001; 2001US-0802706.
XX 14-MAR-2001; 2001US-0808869.
XX 22-MAR-2001; 2001US-0816744.
XX 05-APR-2001; 2001US-0828366.
XX 10-MAY-2001; 2001US-0854208.
XX 10-MAY-2001; 2001US-0854280.
XX 25-MAY-2001; 2001US-0866028.
XX 25-MAY-2001; 2001US-0866034.
XX 25-MAY-2001; 2001WO-US17092.
XX 30-MAY-2001; 2001US-0870574.
XX 30-MAY-2001; 2001WO-US17443.
XX 01-JUN-2001; 2001WO-US17800.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;

PI Stephan JF, Watanabe CK, Williams PM, Wood WT, Ye W;
XX WPI: 2002-090516/12.
DR P-PSDB; ABB84830.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or anglogenetic disorders in a mammal -
XX
XX Claim 2; Fig 27; 565pp; English.
XX
CC ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
CC ABB85003. The PRO proteins and polynucleotides have cardiact, cyostatic,
CC antilangogenic, hypotensive, vulnerary and antiarteriosclerotic
CC activities, and can be used in gene therapy. The PRO polynucleotides,
CC proteins, agonists and antagonists are useful for treating or diagnosing
CC a cardiovascular, endothelial or anglogenetic disorder in a mammal,
CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
CC degeneration, atherosclerosis, hypertension, arterial restenosis,
CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
CC carcinoma) and wound healing. The PRO polynucleotides have applications
CC in molecular biology, including use as hybridisation probes, and in
CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
CC probes used in the exemplification of the present invention.
XX
SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;
Query Match 99.0%; Score 1357.4; DB 24; Length 2005;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATGACAGCAGCTGCACTCATTCACAGGACAGATGAGAACATCCACTATCTCTCAA 60
DB 134 ATGACAGCAGCTGCACTCATTCACAGGACAGATGAGAACATCCACTATCTCTCAA 193
QY 61 GCCACGAGCAGACCTGAAAGACCTGACAGACTTACACAAAGATCCAGAAATAGACA 120
DB 194 GCCACGAGCAGACCTGAAAGACCTGACAGACTTACACAAAGATCCAGAAATAGACA 253
QY 121 GCCATCAAGTTCAACCACTGGAGAGAGCTTCCACTCTTTGAGAGGATATTGGAGC 180
DB 254 GCCATCAAGTTCAACCACTGGAGAGAGCTTCCACTCTTTGAGAGGATATTGGAGC 313
QY 181 ATCATTAGCAATATCATTACACAGCCACACCTCGGAGCGCTGACCAAGCATCTAAT 240
DB 314 ATCATTAGCAATATCATTACACAGCCACACCTCGGAGCGCTGACCAAGCATCTAAT 373
QY 241 GAAGTCAGGACCACTTGACAGATACCTTTACCAACACACAGATATCTGACCTCTTG 300
DB 374 GAAGTCAGGACCACTTGACAGATACCTTTACCAACACACAGATATCTGACCTCTTG 432
QY 301 AATTAATACCTGGCAACATCCGTTGGATTCGTTCTCTCGATGCAACAGATTGG 360
DB 433 AATTAATACCTGGCAACATCCGTTGGATTCGTTCTCTCGATGCAACAGATTGG 492
QY 361 ATGAGGTGAGGTTAGACACTGAGTAGCCAACTTATCAAGTATTTGAGAAATGAAG 420
DB 493 ATGAGGTGAGGTTAGACACTGAGTAGCCAACTTATCAAGTATTTGAGAAATGAAG 552
QY 421 CTAGTAGACTCAAGCATGCTCAGCTCATCAGAAATTTTCAATTAATCTACAAAGTCCACCG 480
DB 553 CTAGTAGACTCAAGCATGCTCAGCTCATCAGAAATTTTCAATTAATCTACAAAGTCCACCG 612
QY 481 GGGCCGAGGGGTCAGAGAGTGACAGAGATCCGAGGAGCCCGTGGCCCACTGGCAAC 540
DB 613 GGGCCGAGGGGTCAGAGAGTGACAGAGATCCGAGGAGCCCGTGGCCCACTGGCAAC 672
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGCCTGACACCACTTGCGCTGGGTGAGAGAGGC 600
DB 673 AAGGACAGAAAGAGAGAGAGGGGAGCCTGACACCACTTGCGCTGGGTGAGAGAGGC 732
QY 601 CCAATTGACACGCTGTGTCGCCCGGAGAGCGTGGCGGCAAAAGATCTAAAGGCTCCAG 660

```

Db 733 CCAATTGGACCGAGCTGTCCCGCCGAGAGACCTGGCGGCAAGATCTAAAGCTCCAG 792
Oy 661 GGGCCCAAAAGGCTCCCGTGGTTCCTCCGGAAGCCCGCCCTCAGAGCCCGCCAGTGGGAGC 720
Db 793 GGGCCCAAAAGGCTCCCGTGGTTCCTCCGGAAGCCCGCCCTCAGAGCCCGCCAGTGGGAGC 852
Oy 721 CCAGGCCCCCGGGCCCAACAGAGGAGGAGCTCCCGGCGCTCAGAGGCGCTCTCTGGC 780
Db 853 CCAGGCCCCCGGGCCCAACAGAGGAGGAGCTCCCGGCGCTCAGAGGCGCTCTCTGGC 912
Oy 781 TTCCAGGAGCTTCAGAGGACCGCTGGGAGACCTGGGCTGACCTCGGGAGCTGCCA 840
Db 913 TTCCAGGAGCTTCAGAGGACCGCTGGGAGACCTGGGCTGACCTCGGGAGCTGCCA 972
Oy 841 GGGCTGGCGTGGGTACCGAGCATGGCCAGGCCCGCCCGGCGCTCTCTGGGCGCA 900
Db 973 GGGTGGCTGGGGGTACCGAGGATGGCCAGGCCCGCCCGGCGCTCTCTGGGCGCA 1032
Oy 901 TCAGAGGCGGTGTGTCCTGGCCCTGGGAGATGAGCCACCCCGGACCGGAGGACAT 960
Db 1033 TCAGAGGCGGTGTGTCCTGGCCCTGGGAGATGAGCCACCCCGGACCGGAGGACAT 1092
Oy 961 GGGTGGCGGCTCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
Db 1093 AGCTGGCCCGGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1152
Oy 1021 GAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGCAAGTCTTCACATCTTTTTCATA 1080
Db 1153 GAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGCAAGTCTTCACATCTTTTTCATA 1212
Oy 1081 AACACTAGAGAGAAACAGCAATGATATAAAACAGATGATGAGAGAGAGACCACTGG 1140
Db 1213 AACACTAGAGAGAGAAACAGCAATGATATAAAACAGATGATGAGAGAGAGACCACTGG 1272
Oy 1141 ATCGGCTTCACAGACTCAGAGCGTGAAGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1200
Db 1273 ATCGGCTTCACAGACTCAGAGCGTGAAGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1332
Oy 1201 GACTACAAAAATTTGGAAGCTGGACAGCGGATTAAGTGGGCTGATGCGCATGGGCGCAGA 1260
Db 1333 GACTACAAAAATTTGGAAGCTGGACAGCGGATTAAGTGGGCTGATGCGCATGGGCGCAGA 1392
Oy 1261 GAAGACTGTGTGGGTGATTATGTTATGCTGGGAGTGAAGATTTTCCAATGTGAAGACGTC 1320
Db 1393 GAAGACTGTGTGGGTGATTATGTTATGCTGGGAGTGAAGATTTTCCAATGTGAAGACGTC 1452
Oy 1321 AATAACTTCATTGGCAAAAGACAGGAGACAGTACTGTCAATGTCATTA 1371
Db 1453 AATAACTTCATTGGCAAAAGACAGGAGACAGTACTGTCAATGTCATTA 1503

RESULT 11
AAH43037
ID AAH43037 standard; cDNA: 2637 BP.
AC AAH43037;
XX
XX
DT 15-OCT-2001 (first entry)
XX
XX Nucleotide sequence of a human scavenger receptor.
KW Human: scavenger receptor; SRCL-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 92..2320
FT /*tag= a
FT /product= "scavenger receptor"
XX

```

```

PN WO200159107-A1.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
XX
XX 14-FEB-2000; 2000JP-0035155.
XX 10-OCT-2000; 2000JP-0309068.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX WPI: 2001-497076/54.
XX P-PSDB: AAG63347.
XX
XX New Scavenger receptor proteins SRCL-PI with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation
XX
XX Claim 6; Page 88-93; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
XX SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;
SQ
Query Match 79.2%; Score 1086.2; DB 22; Length 2637;
Best Local Similarity 87.0%; Pred. No. 1.5e-273;
Matches 1193; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
Oy 1 ATGACAGCGACGCTCACTCACTTCAACAGGCTGAGTGAAGATGACCACTATCTCTCAA 60
Db 947 ATGAAATAGCCACTCACTCACTTCAACAGGCTGAGTGAAGATGACCACTATCTCTCAA 1006
Oy 61 GCCAAGAGCAGAACCTGTAAGACCTGACGACTTACACAAAGATGACAGATAGACA 120
Db 1007 GCCAAGAGCAGAACCTGTAAGACCTGACGACTTACACAAAGATGACAGATAGACA 1066
Oy 121 GCCATCAAGTTCACCACTGAGGAGAGCGCTTCCAGCTCTTTGAGAGGATATTGTGAAC 180
Db 1067 GCTGTCAAGTTCAGCCAACTTGAGGAACGCTTCCAGGCTTTGAGACAGATATTGTGAAC 1126
Oy 181 ATCATTAGCAATATGATTTACACAGCCCAACGCTGCGGAGCGCTACCAAGCAATCTAAAT 240
Db 1127 ATCATTAGCAATATGATTTACACAGCCCAACGCTGCGGAGCGCTACCAAGCAATCTAAAT 1186
Oy 241 GAAGTCAGAGCACCCTTGCACAGATACCTTTACCAAAACACAGATGATGACTCTCTTG 300
Db 1187 GATGTTAGACCAACATGACACAGACACCTTGACACAGACAGAGATGACTGACTCTCTTG 1246
Oy 301 AATAATACCCCTGGCCAACATCCGTTTGATCTCTCTCAGATGCAACAAGATTTG 360
Db 1247 AATAATACCCCTGGCCAACATCCGTTTGATCTCTCTCAGATGCAACAAGATTTG 1306
Oy 361 ATGAGGTGAGGTTGAGACCTGAAGTGAAGCTTATGATTTATGGAAGAAATGAAG 420
Db 1307 ATGAGGTGAGGTTGAGACCTGAAGTGAAGCTTATGATTTATGGAAGAAATGAAG 1366
Oy 421 CTAGTAGACTCCAAAGCATGTCAGCTCATCAAGATTTTACAAATCTACTCAAGTCCACCG 480
Db 1367 CTGTTGACTCCAAAGCATGTCAGCTCATCAAGATTTTACAAATCTACTCAAGTCCACCG 1426
Oy 481 GGGCCCAAGGGGTCCAAGAGTGACAGAGATCCCAAGGAGCCCGTGGCCCAACTGGCAAC 540
Db 1427 GGGCCCAAGGGGTCCAAGAGTGACAGAGATCCCAAGGAGCCCGTGGCCCAACTGGCAAC 1486

```


Db 472 GGCCCTCCTGCTCCAGGAGACTTCAGGGCCACCCTTGGGGAGCCTGGGGTGCCTGGACCT 531
 QY 829 CGGGAGCTCCAGGCTTGCTGGGCTACAGGATGCGCAGGCCCCCAAGGGCCCCCCCCGGC 888
 Db 532 CGGGAGCTCCAGGCTTGCTGGGCTACAGGATGCGCAGGCCCCCAAGGGCCCCCCCCGGC 591
 QY 889 CCTCCTGGCCATCAGAGAGCGGTGTGCCCTTGCCCTCAGAAATGAGCCCAACCCCGGCA 948
 Db 592 CCTCCTGGCCATCAGAGAGCGGTGTGCCCTTGCCCTCAGAAATGAGCCCAACCCCGGCA 651
 QY 949 CCGGAGGAAATGCGTCCCGCCCTCACTGGAAGAACTTACAGACAAATGCTATATTTT 1008
 Db 652 CCGGAGGAAATGCGTCCCGCCCTCACTGGAAGAACTTACAGACAAATGCTATATTTT 711
 QY 1009 TCAGTTGAGAAAGAAATTTTGTGAGATGCAAGCTTTCTGTCGACAAACAGCTTCACAT 1068
 Db 712 TCAGTTGAGAAAGAAATTTTGTGAGATGCAAGCTTTCTGTCGACAAACAGCTTCACAT 771
 QY 1069 CTTGTTTTCATTAACACTAGAGAGAAACAGCAATGATATTAATAAACAAGATGTAGGAGA 1128
 Db 772 CTTGTTTTCATTAACACTAGAGAGAAACAGCAATGATATTAATAAACAAGATGTAGGAGA 831
 QY 1129 GAGAGCCACTGATGCGCTCAGACACTAGAGCGGTGAAGAAATGATGAAAGTGGCTGAT 1188
 Db 832 GAGAGCCACTGATGCGCTCAGACACTAGAGCGGTGAAGAAATGATGAAAGTGGCTGAT 891
 QY 1189 GGGACATCTCCAGACTACAAAAATTTGAAAGCTGAGACCCGATATACGCGGTGCTATGGC 1248
 Db 892 GGGACATCTCCAGACTACAAAAATTTGAAAGCTGAGACCCGATATACGCGGTGCTATGGC 951
 QY 1249 CATGGGCCAGGAGAACTGTCTGGGTGATTTATGTGGGCACTGAGAACATTTCCAA 1308
 Db 952 CATGGGCCAGGAGAACTGTCTGGGTGATTTATGTGGGCACTGAGAACATTTCCAA 1011
 QY 1309 TGTGAAGACGTAATTAATTTGCGAAAGAAAGACAGGAGACACT 1355
 Db 1012 TGTGAAGACGTAATTAATTTGCGAAAGAAAGACAGGAGACACT 1058

RESULT 13

AAH43054 standard; DNA; 2256 BP.

XX AAH43054;
 AC AAH43054;
 XX
 DT 15-OCT-2001 (first entry)
 XX
 DE Nucleotide sequence of a human scavenger receptor.
 XX
 KW Human: scavenger receptor; SRCL-P1; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KM low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 74..1936
 FT /tag= a
 FT /product= "scavenger receptor"
 PN WO200159107-A1.
 PD 16-AUG-2001.
 XX
 PF 08-FEB-2001; 2001WO-JP00874.
 XX
 PR 14-FEB-2000; 2000JP-0035155.
 PR 10-OCT-2000; 2000JP-0309068.
 XX
 PA (FUSO) FUSO PHARM IND LTD.
 XX
 PI Wakamiya N;
 XX

DR WPI: 2001-497076/54.
 DR P-PSDB: AAG63350.
 XX
 PT New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 PT useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation
 XX
 PS Claim 4; Page 105-109; 118pp; Japanese.
 CC The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.
 SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other:

Query Match 45.0%; Score 617; DB 22; Length 2256;

Best Local Similarity 72.9%; Pred. No. 6, 2e-151;

Matches 999; Conservative 0; Mismatches 0; Indels 372; Gaps 1;

QY 1 ATGACAGCAGCTCAACTCATTCACAGGTACAGATGAGAAATCAACCACTATCTCTCAA 60
 Db 929 ATGACAGCAGCTCAACTCATTCACAGGTACAGATGAGAAATCAACCACTATCTCTCAA 988
 QY 61 GCCACAGCAGACCTGAAAGACCTGAGAGACTTACACAAAGATCAGAGATAGAGAA 120
 Db 989 GCCACAGCAGACCTGAAAGACCTGAGAGACTTACACAAAGATCAGAGATAGAGAA 1048
 QY 121 GCCATCAAGTTCAACCACTGAGAGAACCTTCCACTCTTTAGACGATATTTGGAAC 180
 Db 1049 GCCATCAAGTTCAACCACTGAGAGAACCTTCCACTCTTTAGAGCGATATTTGGAAC 1108
 QY 181 ATCATTAGCAATATTCAGTTACACAGCCCAACCTCGGAGCGGTGACCAATCTCAAT 240
 Db 1109 ATCATTAGCAATATTCAGTTACACAGCCCAACCTCGGAGCGGTGACCAATCTCAAT 1168
 QY 241 GAAGTCAGAGACCACTGACAGATACCTTACCAAAACAGACAGATATGACCTCTTG 300
 Db 1169 GAAGTCAGAGACCACTGACAGATACCTTACCAAAACAGACAGATATGACCTCTTG 1228
 QY 301 AATAATACCTGCGCAACATCCGTTTGATTCCTTCTCTCAGATGCAACAAGATTGG 360
 Db 1229 AATAATACCTGCGCAACATCCGTTTGATTCCTTCTCTCAGATGCAACAAGATTGG 1288
 QY 361 ATGAGTCGAGGTTACACTGAGATGAGCAACTTATTCAGTGTGTAAGAAATGAAG 420
 Db 1289 ATGAGTCGAGGTTACACTGAGATGAGCAACTTATTCAGTGTGTAAGAAATGAAG 1348
 QY 421 CTAGTAGACTCCAGAGATGTCAGTCATCAAGATTTTACATTAAGATGATCCACCG 480
 Db 1349 CTAGTAGACTCCAGAGATGTCAGTCATCAAGATTTTACATTAAGATGATCCACCG 1408
 QY 481 GGGCCCAAGGGGTCACAGAGTGCAGAGATCCAGAGAACCCCTGCGCAACTGCAAC 540
 Db 1409 GGGCCCAAGGGGTCACAGAGTGCAGAGATCCAGAGAACCCCTGCGCAACTGCAAC 1468
 QY 541 AAGGACAGAAAGAGAGAAAGGGGAGCCTGGAGCAACCTGGCCCTGGGTAGAGAGG 600
 Db 1469 AAGGACAGAAAGAGAGAAAGGGGAGCCTGGAGCAACCTGGCCCTGGGTAGAGAGG 1516
 QY 601 CCAATTGACACAGCTGTCCCCCGGAGAGCGTGGCGCAAAAGATCTAAGGCTCCAG 660
 Db 1517 ----- 1516
 QY 661 GGGCCCAAGAGCTCCCGTGGTTCCCTGGGAGAGCCCGCCCTAGAGGCCCACTGGGAGC 720
 Db 1517 ----- 1516

```

QY 721 CCAAGCCCCCGGACCACAGGCAAGAGGAGCTCCCGCCCTCAGGGCCCTCCTGGC 780
Db 1517 ----- 1516
QY 781 TTCAGGAGACTTCAGGGACCGCTTGGGAGCCCTGGGCTGACCTCGGGACACTCCA 840
Db 1517 ----- 1516
QY 841 GCGTTGCGTGGGTACCAGGCAATGCCAGGCCCAAGGGCCCCCGCCCTCCTGGCCCA 900
Db 1517 ----- 1516
QY 901 TCAGAGAGCGGTGGTCCCTCGGCCCTGACAGATGAGCCAAACCCCGGACCGGAGAGCAAT 960
Db 1517 ----- 1516
QY 961 GCGTCCCGGCTCAGTGGAGAACTTCAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
Db 1517 GCGTCCCGGCTCAGTGGAGAACTTCAGACAAATGCTACTATTTTTCAGTTGAGAAA 1576
QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGCTTTCACATCTTGTTCATA 1080
Db 1577 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGCTTTCACATCTTGTTCATA 1636
QY 1081 AACACTGAGAGAGACAGCAATGATTAATAAAGACAGATGTTAGGAGAGAGAGCACTGG 1140
Db 1637 AACACTGAGAGAGAGCAATGATTAATAAAGACAGATGTTAGGAGAGAGAGCACTGG 1696
QY 1141 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAATGAGTGGCTGATGGGACATCTCCA 1200
Db 1697 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAATGAGTGGCTGATGGGACATCTCCA 1756
QY 1201 GACTACAAAAATTGGAAAGCTGACAGCCGAGTAACTGGGGTCATGGCCATGGGCCAGGA 1260
Db 1757 GACTACAAAAATTGGAAAGCTGACAGCCGAGTAACTGGGGTCATGGGCCATGGGCCAGGA 1816
QY 1261 GAAGACTGCTGGTGGATTTATGCTGGGACATGAGAACATTTCCATATGGAAGAGCTC 1320
Db 1817 GAAGACTGCTGGTGGATTTATGCTGGGACATGAGAACATTTCCATATGGAAGAGCTC 1876
QY 1321 AATTAATCTCAATTTGCGAAAAAGACAGAGAGACAGTACTGATCTGCAATTA 1371
Db 1877 AATTAATCTCAATTTGCGAAAAAGACAGAGAGACAGTACTGATCTGCAATTA 1927

RESULT 14
AAx27858
ID AAx27858 standard; DNA; 3685 BP.
XX
XX AAx27858:
XX
XX 02-JUN-1999 (first entry)
XX
XX
XX Human CSR3 protein coding sequence.
XX
XX Cellular stress response protein: CSR1; CSR2; CSR3; human; macrophage;
XX scavenger receptor protein; intracellular stress; arteriosclerosis;
XX diabetic circulatory obstruction; microbial infection; ss.
XX
XX Homo sapiens.
XX
XX WO990159-A1.
XX
XX 25-FEB-1999.
XX
XX 12-AUG-1998: 98WO-JP03602.
XX
XX 30-JUL-1998: 98JP-0230121.
XX
XX 13-AUG-1997: 97JP-0233396.
XX
XX (NIBS) JAPAN TOBACCO INC.
XX
XX Nakamura Y, Tokyo T.
PI

```

```

XX
XX WPI: 1999-181032/15.
DR P-PSDB: AAY00994.
XX
XX Scavenger receptor proteins - for treatment and diagnosis of
PT disorders involving cell stress
XX
XX Claim 6: Page 142-150; 175pp; Japanese.
XX
XX This sequence encodes the human cellular stress response 3 (CSR3) protein
XX of the invention. The CSR proteins are macrophage scavenger receptor
XX proteins. The CSR proteins can be used in the treatment, gene therapy
XX and diagnosis of diseases in which intracellular stress is important,
XX such as arteriosclerosis, diabetic circulatory obstruction, and microbial
XX infection. Expression of the proteins is induced in vivo in response to
XX intracellular stress, and inhibits cell death as a result of such stress.
XX
XX Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other:
XX
XX Query Match 17.2%; Score 235.6; DB 20; Length 3685;
XX Best Local Similarity 53.9%; Pred. No. 3.7e-51;
XX Matches 484; Conservative 0; Mismatches 414; Indels 0; Gaps 0;
XX
QY 9 CCAAGCTCACTCATTCACAGTGCAGATGAGAGAACATTCACACTATCTCTAACCCACGA 68
Db 1056 CCAAGTCACTGAGCTTGCAGCTGAGTAACTCTCTCTCTCTGATGACACGA 1115
QY 69 GCAGAACCTGGAAGACCTCGAGACTTACACAAAGATGCAGAGATTAAGACCCATCAA 128
Db 1116 AGAGAACATCATGATCTTACATGATACCTACCTACCTACCTACCTGATGAGAG 1175
QY 129 GTTCAACCAATGAGAGACCTTCCAGCTTTGAGAGCGAATTTGTGACATCATTAG 188
Db 1176 GTTGAAGTCTCTGGAAGAGAGCATGCTTCTCAGAGATTAAGATGACATCTTAC 1235
QY 189 CAATATCAGTTACAGACCCACCGAGCGAGCGTACACAGCAATCTAATGAAGCAG 248
Db 1236 CAATATCAGTTACAGACCCACCGAGCGAGCGTACACAGCAATCTAATGAAGCAG 1295
QY 249 GACCACTTGCACAGATACCTTACCAAAACACAGATGATCTGATCTTGTATTAATAC 308
Db 1296 GCTCTCTGAGCGCTGAGCTTCCACACCGAGAGAGCTCTACTACTGAAACAAAGTC 1355
QY 309 CTTGGCCACATCCTGTTTGGATTTCTGTTCTCAGAGATGCAACAAAGATTTGATGAGTC 368
Db 1356 TGTCTCTCATCATGCTGGGACACAGACCTCTCGGAGAGCGTTCAGCTGCTCAGTCG 1415
QY 369 GAGTTTGAACAGTGAAGTACCACTATCATGATTAATGAAGAAATGAAGCTATGAG 428
Db 1416 CCGGCTGGAGCTCAAGCTCCGGAACCTCTCATGATCTGAGAGAGATGAAGCACTGGA 1475
QY 429 CTCCAAGCATGTCAGCTCAAGATTTTACAAATCTCAAGAGTCCACCGGCCACAG 488
Db 1476 CACACAGCATGGAAGAAATCTTGGCAATGTCACATCTCTGAGAGATGAGAGAGAG 1535
QY 489 GGGTCCAGAGGTGACAGAGATCCAGGAGACCCCTGCGCCCACTGGCAACAGGAGACA 548
Db 1536 AGGACCAAGAGATTAAGAGAGATATGGCGCTGTAAGAGCGCTGTTGGCGGAGGCC 1595
QY 549 GAAAGAGAGAGAGGAGAGCTGAGACCACTGGCCCTGCGCGGTGAGAGAGGCCAATGG 608
Db 1596 GAAAGAGAGAGAGGAGAGCTGAGACCACTGGCCCTGAGAGAGGCCAATGGAGCAACCTGG 1655
QY 609 ACCAGCTGTCCTCCCGGAGAGCGTGGCGCAAGATCTTAAGGCTCCAGAGGCCCAA 668
Db 1656 AGAGGCGGCGCTGCTGGGAGAAAGGGGCGCTGTTGGCCCTCGAGGGTTCCAGAGCTCAA 1715
QY 669 AGGCTCCCGTGTCCCTGGAAGCCCGGCTCAGGGGCCAGTGGGGAGCCAGGCC 728
Db 1716 AGGCTCAAGAGGAGGCTTTGAACTGGAAGGCGGAGAGAGAGAGGCCCAAGAGGGA 1775
QY 729 CCGGGGCCACACAGGAGAGGAGTCCCGGCGCTCAGAGGCGCTTCTGCTTCCAGGG 788
Db 729 CCGGGGCCACACAGGAGAGGAGTCCCGGCGCTCAGAGGCGCTTCTGCTTCCAGGG 788

```

Db 1776 CATAGGGCCCCAGAGGCCAGAAAGGGCCCCCGGGGTCTCCAGAGGCCCTTCAGAGGCTTCAGAGG 1835

Qy 789 ACTTCAGGGGACACGTTGGGGAGACCTGGGGTGCCTGGACCTCGGGGACTCCAGGCTTGGCC 848

Db 1836 AAAACCGCGAAATTGGACGGGAAAGACAGGGCTACACAGGCCAGCGGGGGCCATGGGGCTTAA 1895

Qy 849 TGGGGTACCGAGCATGCGCAGCGCCCAAGGGGCCCGCCGGCCCTCGCTGGCGCCATMGAGGA 906

Db 1896 GGGTACACCGAGGATCCAGGGTCCCCCTGGTCTCCGGGGGCTCTCCAGGTCCACCAAGGA 1953

RESULT 15

ID AAX27856

AXX27856 standard; DNA; 3810 BP.

AXX27856;

DT 02-JUN-1999 (first entry)

XX Human CSRI protein coding sequence.

DE Cellular stress response protein; CSRI; CSR2; CSR3; human; macrophage;

KW scavenger receptor protein; intracellular stress; arteriosclerosis;

KW diabetic circulatory obstruction; microbial infection; ss.

XX OS Homo sapiens.

XX PN W09909159-A1.

XX PD 25-FEB-1999.

XX PF 12-AUG-1998; 98WO-JP03602.

XX PR 30-JUL-1998; 98JP-0230121.

XX PR 13-AUG-1997; 97JP-0233396.

XX PA (NISB) JAPAN TOBACCO INC.

XX PI Nakamura Y, Tokyo T;

XX DR WPI; 1999-181032/15.

XX P-PSDB; AAY00992.

PT Scavenger receptor proteins - for treatment and diagnosis of

PT disorders involving cell stress

PS Claim 4; Page 119-127; 175pp; Japanese.

CC This sequence encodes the human cellular stress response 1 (CSRI) protein

CC of the invention. The CSR proteins are macrophage scavenger receptor

CC proteins. The CSR proteins can be used in the treatment, gene therapy

CC and diagnosis of diseases in which intracellular stress is important,

CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial

CC infection. Expression of the proteins is induced in vivo in response to

CC intracellular stress, and inhibits cell death as a result of such stress.

XX

SO Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;

Query Match 17.2%; Score 235.6; DB 20; Length 3810;

Best Local Similarity 53.9%; Pred. Mismatches 414; Indels 0; Gaps 0

Matches 484; Conservative 0;

Qy 9 CCAGCTCAACTCATTTCACAGGTCCAGATGGAGAAACATCACACTATCTCAAGCCAAAGCA 68

Db 1181 CCAGGTCAATGGGCTTGCACGTGCACGTGGATTAACATCTGCTCTTGATGATCACCAGCA 1240

Qy 69 GCAGAACTGAAAGACCTGCGAGACTTACACAAAGATGCGAGAGATAGAACAGCCATCAA 128

Db 1241 AAGAGAACATGATGATCTTCAGTACCATACCCACTGCGCCGAAACCGCATGTGGAGAG 1300

Qy 129 GTTCAACCAACATGGGAGGAGGCTTCCAGCTCTTTGGAGAGGGAATTGTGAACATCTTAG 188

Db 1301 GTTGTGCTCTGGAGAGGCGATGGCTTCTCAACGAGATTGAATGGACCATCTTTCAC 1360

QY	189	CAAAATGATTAACAAAGCCCAACCTGGGAGAGCGTACACCAATCTAAATGAAGTAC	248
Db	1361	CAACATCAATATGCCACCGACACACCACTGCAACAGCATGCTCAAGTACGTGATGACGTGCG	1420
QY	249	GACCCTTGGACAGATACCTTTACCAAAACACACAGATGATCGACCTCTTGAATATATAC	308
Db	1421	GCCTCCCTGCACGCTGGGCTTCCACACCCATCGCCGAGAGCTCTACTACCTGAACAAATGC	1480
QY	309	CCCTGGCAACATCCGTTTGGATTTCTGTCTTCCTCGAGGATGCAACAAAGTTGATGAGGTC	368
Db	1481	TGTCTTCATATGCTGGGACACACACACCTGCTCGGGAGGCTTACAGCTGTCTCATGTGC	1540
QY	369	GAGGTTAGACATCGAAGTATGACCACTTATCAGTATGATTAAGAAAGAAATGAAGTATGAGA	428
Db	1541	CCGGCTGACCTCAAGCTCCGGAACCTCTGCATGATCGTGGAGGATGAAAGGACGTGA	1600
QY	429	CTCCACACATGTGTACGTCTACTCAAGATTTTACATACTACAAAGTCCACGGGCCCCAG	488
Db	1601	CACACACATGAGGAAATCTTTGGCAATGTACACATCTTACAGAGGTGCCCGCGGCTCC	1660
QY	489	GGGTCAAGAGGTACAGAGAGATCCGAGGACCCCTGGCCCACTGGCCCAACAAGGACAA	548
Db	1661	AGGACCAAGAGGANTTCAAAAGATATGTGGCGGTGAAGGGCCCTGTGTGGCGGACAGGGCC	1720
QY	549	GAAGAGAGAGAGGGGAGGCTGTGACCACTGGCCCTCGGGGTGAGAGAGGCCCAATTGG	608
Db	1721	GAAGAGGAGACCCGGCTTCTTTGGGGCCCCCTGGGACCCCAAGGCTCTCCAGGGGCAACCTGG	1780
QY	609	ACCAAGGTGTCCCCCGGAGAGGCTGGCGGCAAAAGATTTAAAGCTCCCAAGGGCCCCAA	668
Db	1781	AGAGGCGGGGCTGTGTGGAGAAAGGGGCCCTGTGTGGCCCTCGAGGGTTTCCAGGCTCTAA	1840
QY	669	AGGCTCCCGTGTCCCTCGGGAAGCCCGAGCCCTCAGGGGCCCAATGTGGGAGACCAAGGCC	728
Db	1841	AGGCTCAAAAGGAGCTTTGAACTGGAAGGGGCCGAGAGGACACCAAGGCCCAAAAGGGGA	1900
QY	729	CCCGGGGCAACAGGCAAAAGGAGATCCCGGGCCCTCGAGGGCCCTCCGTGGGCTTCAAGG	788
Db	1901	CATTAGGGCCCCCAGGCCCAAGAGGCCCCCGGGGTTTCCAGGGGCCCTCAGGGGCTCAAGGG	1960
QY	789	ACTTCAGGGACCGTTGGGGAGGCTTGGGGTGCCTGGACCTCGGGGACCTGCAAGGCTTGCC	848
Db	1961	AAAACCGGGAATTTCAGGGAAGACAGGATCACAGGCCACGCGGGGGGACATATGGGACTTAA	2020
QY	849	TGGGGTACCAAGGATGCGCAGGGCCCAAGGGCCCCCGGGCCCTCCGTGGGCTTCAACAGGA	906
Db	2021	GGGTGAACCAAGGATTCAGGGTCCCTCCCTGTCTTCCCGGGGCTTCCAGGTTCAACAGGA	2078

Search completed: March 21, 2003, 02:19:04
Job time : 298.589 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 02:00:05 ; Search time 48,326 Seconds
(without alignments)
8700.356 Million cell updates/sec

Title: US-09-763-712A-1_COPY_325_1695

Perfect score: 1371

Sequence: 1 atgaacagcagcgcacac...caactatcgtcatcgaatta 1371

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued Patents_NA.*
- 2: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/6C.COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.6	8.9	810	1	US-08-642-255-60
2	118	8.6	756	1	US-08-642-255-50
3	113.6	8.3	3181	1	US-08-655-086-1
4	112.8	8.2	506	1	US-08-253-155A-61
5	106	7.7	432	1	US-08-642-255-48
6	106	7.7	1608	4	US-09-028-348-19
7	105.4	7.7	503	1	US-08-253-155A-20
8	100	7.3	5102	1	US-08-494-168-1
9	99.6	7.3	1416	2	US-07-621-091G-1
10	99.6	7.3	1416	2	US-08-399-889-1
11	99.6	7.3	1416	2	US-09-167-364-1
12	99.6	7.3	1416	4	US-09-439-897-1
13	98.4	7.3	1868	1	US-08-392-367B-1
14	98.4	7.3	1868	3	US-08-893-467A-1
15	99.4	7.3	2543	3	US-08-555-669-11
16	99.4	7.3	2543	3	US-09-073-663-11
17	97.6	7.1	4359	4	US-09-484-970B-4
18	96.2	7.0	1572	4	US-09-297-268-39
19	95.2	6.9	1560	2	US-08-794-795-5
20	95.2	6.9	1560	4	US-09-249-200-5
21	95.2	6.9	1703	2	US-08-794-795-1
22	95.2	6.9	1703	4	US-09-249-200-1
23	89	6.5	1881	4	US-09-029-348-20
24	77	5.6	1560	4	US-09-453-702B-264
25	77	5.6	61663	4	US-09-453-702B-62
26	76.8	5.6	5552	3	US-08-155-888-1
27	74.6	5.4	38584	4	US-09-453-702B-50

28	73	5.3	503	4	US-09-297-269-40	Sequence 40, Appl
29	72.4	5.3	4031	1	US-08-159-784-1	Sequence 1, Appl
30	72.2	5.3	9827	4	US-09-453-702B-66	Sequence 66, Appl
31	71.4	5.2	3922	4	US-09-404-879A-372	Sequence 372, App
32	70	5.1	513	4	US-09-134-001C-647	Sequence 647, App
33	70	5.1	585	4	US-09-134-001C-726	Sequence 726, App
34	70	5.1	810	4	US-09-134-001C-624	Sequence 624, App
35	70	5.1	3552	4	US-09-134-001C-693	Sequence 693, App
36	69	5.0	1347	4	US-09-140-804-1	Sequence 1, Appl
37	68.6	5.0	45175	4	US-09-453-702B-116	Sequence 116, App
38	67.8	4.9	3404	4	US-09-453-702B-94	Sequence 94, Appl
39	67.8	4.9	48908	4	US-09-453-702B-137	Sequence 137, App
40	67.2	4.9	729	4	US-09-140-804-10	Sequence 10, Appl
41	67.2	4.9	226	1	US-08-175-155-63	Sequence 63, Appl
42	67	4.9	226	1	US-08-477-509B-88	Sequence 88, Appl
43	67	4.9	226	1	US-08-642-255-96	Sequence 96, Appl
44	67	4.9	226	1	US-08-397-633A-16	Sequence 16, Appl
45	67	4.9	226	2	US-08-707-237A-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-08-642-255-60
Sequence 60, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"
US-08-642-255-60
Query Match 8.9%; Score 121.6; DB 1; Length 810;
Best Local Similarity 53.9%; Pred No. 4.1e-23;
Matches 250; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
QY 465 ACTACAAGTCCACCGCGGCCAGGGGTCCAGAGGTGACGAGATCCAGGAGCCCC 524
|| || ||| |||| ||||| ||||| || ||||| ||

Db	210	ACCAACGGGTGCACCTTGCCAGCGGGGTCCGGCTCGATCTAGAGGTGAACCCAGGACCGCC	269
Oy	525	TGCCCCAATCTGGCAACAAGGAGACGAAGAAGAGAGAGGGGAGCGCTTGACCTACCTGACC	584
Db	270	TGCTGCTCCGGAGACTCGTAGGACCACCGGGGTAGCGCGATCCGGAGACCCGGGTGC	329
Oy	585	TGCGGGTGAAGAGAGGCCCAATTGGACCACTGGTCCCGCCGGAGAGCGCTGGCGCAAGG	644
Db	330	ACCTGGGCCACAGCGGTCCGCGCTCGATCTAGAGGTGACCCAGACCGCGCTGTGCTCCGGG	389
Oy	645	ATCTAAGGCTTCCAGAGGCCCCCAAAAGGCTCCCGTGGTCCCTCTGGAGAGCGCGGCTCTCA	704
Db	390	ACCTGACGCCCCACCGGGTACCGTGGCGATCCGGGACCAACCGGTGTACACTTGACCCAGC	449
Oy	705	GGGGCCCAAGTGGGGACCCAGGCGCCCCGGGGGCCACAGGCAAGAGGGAATCCCCGGGCC	764
Db	450	GGGTCCCGCTGGATCTAGAGGTGTGCCAGAGACCGCTCGTGGCTCCGGGACTTCAGAGCC	509
Oy	765	TCAGAGGCCCTCTGGCTTCCAGGGACCTTCAGGGACACCTGGGGAGCGCTGGGGGCTCGG	824
Db	510	ACCGGGTACCGGTGGCGATCCGGGAGCACCGGGGTGCACTTGGCCCAACGGGGTCCGGCTGG	569
Oy	825	ACCTGGGGAGCTGCAGAGCTTGCCTGGGGTACAGAGCATGCAGAGGCCCAAGGAGCCCCC	884
Db	570	ATTCTAGAGGTACCCAGGACCGCTGTGTCTCCGGGACTGTGCAAGGCCCAACCGGGTACCGG	629
Oy	885	CGGCGCTCTGGCCATTCAGAGGGGTGGTGGTCCCTGGACCTGCG	928
Db	630	TGGCGATTCGGGAGCCACCGGGGTGCACCTTGCGCCAGCGGGTCCCG	673

```

1 RESULT 2
2 US-08-642-255-50
3 Sequence 50, Application US/08642255
4 Patent No. 5773249
5
6 GENERAL INFORMATION:
7 APPLICANT: CAPPELLO, Joseph
8 APPLICANT: FERRARI, Franco A.
9 TITLE OF INVENTION: High Molecular Weight Collagen-Like
10 TITLE OF INVENTION: Protein Polymers
11 NUMBER OF SEQUENCES: 135
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: FLEHR, HORBACH, TEST, ALBRITTON & HERBERT
14 STREET: 4 Embarcadero Center, Suite 3400
15 CITY: San Francisco
16 STATE: California
17 COUNTRY: USA
18 ZIP: 94111-4187
19
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Floppy disk
22 COMPUTER: IBM PC compatible
23 OPERATING SYSTEM: PC-DOS/MS-DOS
24 SOFTWARE: PatentIn Release #1.0, Version #1.30
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/08/642,255
27 FILING DATE:
28
29 CLASSIFICATION: 435
30 ATTORNEY/AGENT INFORMATION:
31 NAME: ROWLAND, Berttram I.
32 REGISTRATION NUMBER: 20, 015
33 REFERENCE/DOCKET NUMBER: A55556-3/BIIR
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (415) 494-8700
36 TELEFAX: (415) 494-8771
37 TELEX: 910 277299 FHT UR
38
39 INFORMATION FOR SEQ ID NO: 50:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 756 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: double
44 TOPOLOGY: linear
45
46 MOLECULE TYPE: other nucleic acid
47 DESCRIPTION: /desc = "synthetic"
48
49
50

```

Query Match	8.6%;	Score 118;	DB 1;	Length 756;
Best Local Similarity	54.3%;	Prod. No. 3.6e-22;		
Matches 238;	Conservative 0;	Mismatches 200;	Indels 0;	Gaps 0;
QY 472	GATCCACCGGGGCCCCAGGGGGTCCCAAGAGGTGACAGAGGATCTCCACGAGGACCCCTGAGGCCA	531		
Db 289	GATCCACCGGGTGTCTCCGGGACCTTCAGGCCCGCCAGGTGCGCTGAGACCGGCTGTCCA	348		
QY 532	ACTGGCAACAAGGGACAGAAAGAGAGAAAGGAGGAGGAGCCCTGACCACTGACCTCGGGT	591		
Db 349	CCGGGTGCTCCGGGACCTGACAGGCCCGCCAGGTGAGGCTGAGACCGGCTGTACACCGGTT	408		
QY 552	GAGAGAGGCCCAATTGACACAGCTGTGTCCCGCCGAGAGAGCTGTGCGCGCAAGAGATCTAAA	651		
Db 409	GCTCCGGGACCTGACAGGCCCGCCAGGTGCGCTGAGCCGCGGTGCTCCACCGGGTGTCCG	468		
QY 652	GAGTCCCAAGGGGCCCAAGAGGCTCCGCTGTTCCTCCCTGGGAACCCCGGCCCTCAGAGGCCCC	711		
Db 469	GGACCTGCAGAGGCCGCCACAGGTGCGCTGTGACCGGCTGTGCCACCGGGTGTCCCGGAGCTT	528		
QY 712	AGTGGGACACAGAGGCCCGCCGAGCCACACAGGCAAGAGAGGATCCCTCGGACCTAGAGG	771		
Db 529	GCAGGCCCGCCACAGGTGCGCTGTGACCGGCTGTCTCACCGGCTGTCTCCGGGACCTGACAGCC	588		
QY 772	CCTCTGTGCTTCAGAGGACTTCAGGGCAACCGTTTGGGAGGCCGTGGGGTGTGCTGAGACTCGG	831		
Db 589	CCGCCAGGTGGCGCTTGCAGCCGCGCTGTCCACCGGGTGTCTCCGGGACCTGACAGGCCGCCA	648		
QY 832	GGACTGCCAAGGCTTTCCTGTGGGGTACCAAGCATGCCAAGGCCCAAGGGCCCCCGGGCCTT	891		
Db 649	GGTGGCGCTTGGACCGGCTGTGTCACCGGCTGTCTCCGGGACCTGACAGGCCCGCCAGGTGCC	708		
QY 892	CCGTGGCCCATCAGAGGCG 909			
Db 709	CATGGCCCAAGGACCG 726			

US-08-655-086-1
Sequence 1, Application US/08655086
Patent No. 5821089
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELIOTT A.
APPLICANT: BUECHTER, DOUGLAS
APPLICANT: ZHANG, GUANGHUI
APPLICANT: CONNOLLY, KEVIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILMORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: US
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,086
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S.
REGISTRATION NUMBER: 32,063
REFERENCE/DOCKET NUMBER: 203-1632
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-228-8484
TELEFAX: 516-228-8516
INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
US-08-655-086-1

```

```

Query Match      8.3%; Score 113.6; DB 1; Length 3181;
Best Local Similarity 52.5%; Pred. No. 1e-20;
Matches 248; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

```

```

QY 442 CAGCTATCAAGAAATTTTACATACACTACAGAGTCCACCGGGCCCGAGGGCTCAAGAGGT 501
DB 104 CTGGTCCCAAGAGCTTCCCAAGGCTTCCAGAGTCCCGCTGGAGAGCTGCGACGCTGGA 163
QY 502 GACAGAGATCCAGAGGAGCCCGCTGCGCCCACTGGCAACAGAGGACAGAAAGAGAGAAG 561
DB 164 GCTTACAGTCCCATGGGTCCCGAGAGTCCCGCAGGTCCCGCTGAAAGAAATGAGATGAT 223
QY 562 GGGGAGCCCTGAGACACCTGGCCCTGGGGTGAAGAGGCCCAATTGGACAGCTGTGCC 621
DB 224 GGGGAGGCTGAGAAACCTGTGCTCTGTGAGCGTGGGCTCTCTGGGCTCAGAGGTGT 283
QY 622 CCCGAGAGCTGGCGGCAAGATCTAAAGCTCCAGAGGCCCAAGAGCTCCCGTGT 681
DB 284 CGAGATTCGCCGGAACAGACTGGCTCCCTGGAATGAAGGACACAGAGTTTCACTGT 343
QY 682 TCCCTGGAGAGCCCGGCTCAGGGCCCACTGGGGAGCCAGGCCCGCGGCCACCA 741
DB 344 TTGATGTGGTCCCAAGGAAATGTGTGCTGCTGCTAAAGGTGAGCTGGCAGCCCT 403
QY 742 GGCAAGAGGAGACTCCCGGCTCAGGGCCCTCCGCTTCCAGGAGTTTCAAGGCAC 801
DB 404 GTGAAATGAGAGCTCTGTGTCAGATGGGCCCGCTGCTGCTGAGAGAGGTGCG 463
QY 802 GTTGGGAGAGCTGGGTGCTGAGACTCGGGAGACTCCAGGCTTGGGGTACAGAGC 861
DB 464 CCTGAGAGCCCTGCGCTGTGTGTGTGGAATGATGATGCTGTGTGTGTGTGTGTGTGT 523
QY 862 ATCCAGAGCCCGAGAGGCCCGGCTGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 913
DB 524 CCCCTGTGCTCCAGCGGCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 575

```

```

RESULT 4
US-08-235-155A-61/c
; Sequence 61, Application US/08253155A
; Patent No. 5691147

```

```

; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709

```

```

; REFERENCE/DOCKET NUMBER: MIT-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-253-155A-61

```

```

Query Match      8.2%; Score 112.8; DB 1; Length 506;
Best Local Similarity 53.8%; Pred. No. 7.5e-21;
Matches 254; Conservative 0; Mismatches 217; Indels 1; Gaps 1;

```

```

QY 182 TCATTAGCAATATCAGTTTACAGAGCCACCACTGGAGAGCTG-ACCAGCAATCAAT 240
DB 506 TCTTACCAACATCATATTCCACGACACAGTGTGACAAACATGCTCCAAAGTACCTGGAT 447
QY 241 GAAGTCAGACCACTTGCACAGATPACCTTACCAACACACAGATGATGTGACTGCTTG 300
DB 446 GACGTTCGGCTCTCCGCAAGCTGGGCTTCCACACCCATGCGAGAGGCTTACTACTG 387
QY 301 AATTAATACCTGGCCCAACATCCGTTTGATTTGTTTCTCAGAGTGCACAAAGTTTG 360
DB 386 AACAGTGTGTCTCCATCATGTCTGGGACACACAGACTCTCTCGGAGCGCTTCAAGCTG 327
QY 361 ATGAGTCTGAGGTATGACACTGAAGTACCACTTATCATGATTTATGAAGAAATGAG 420
DB 326 CTGAGTCCCGCTGAGCTCAAGCTCGGAACTCTCATGATGCTGAGAGAGATGAG 267
QY 421 CTAGTACTCCAGATGTGTAAGCTCATCAAGATTTTACAACTACAGATTCACAGC 480
DB 266 GCAGTGGACACAGCATGAGAAATCTTCGCAATGTCACATCTTACAGAGTGTGCC 207
QY 481 GGGCCGAGGGTCCAGAGGTGACAGAGATCCAGAGGCCCGCTGGGCCCACTGGCAG 540
DB 206 GGCTCTCCAGAGACAGAGATTCACAGAGATGAGGAGTGAAGAGGCTGTGTGGGCG 147
QY 541 AAGGAGAGAAAGAGAGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 146 AGAGGCGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 87
QY 601 CCAATTGACACAGCTGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
DB 86 CAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 35

```

```

RESULT 5
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249

```

```

; GENERAL INFORMATION:
; APPLICANT: Cappelletto, Joseph
; APPLICANT: Ferrarri, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHER, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```


LENGTH: 503 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-08-253-155A-20

Query Match 7.7%; Score 105.4; DB 1; Length 503;
 Best Local Similarity 52.9%; Pred. No. 7e-19;
 Matches 226; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

63 CACGACGAGAACTGAAAGCCTCGAGAGCTTACCAAAAGATGCAAGATAGAAACAC 122
 15 CCACGAGAGAACATGATGATCTTACATACCTACCTACGCCAAGACCGCACTGT 74
 123 CATCAGTTTCAACCACTGAGAGAGAGCTTCCAGCTCTTTGAGAGGATTTGTGAAC 182
 75 GGAGAGTTTGTGCTCTGTGTAGAGAGCATGCTTCTCAGAGATTTGAATTTGGCAC 134
 183 CATTAAGCATATCAGTTACACAGCCACCTGCGAGAGCTGACCAAGATTAATGA 242
 135 CTTACACCAATCATATGCGACGACGACGACGACGACGACGATGCTATGATGATGA 194
 243 AGTCAGAGCACTTGCACAGATATACCTTACCAAAACAGATGATCTGACCTCTTGA 302
 195 CGTGGGCTCTCTGACAGCTGGGCTTCCACACCATGCGGAGAGCTCTACTACTGA 254
 303 TAATACCTGCGCAATCCGTTTGGATTTCTTCTCAGAGATGCAACAAAGATTGAT 362
 255 CAAGTCTGTCTCATCATGCTGTGGGACCAACACCTGCTCCGCGAGGCTTACAGCT 314
 363 GAGTCGAGAGTTAGACACTGAGTAGACCACTTATCAGTATTATGAGAAATGAAGCT 422
 315 CAGTGGCCGCTGAGACCTTCAAGCTCCGGAACCTTCTCATGATGCTGAGAGATGAAG 374
 423 AGTAGATTCACAGCATGTGCTCATCAAGATTTTACAAATATACAAAGTCCACAGG 482
 375 AGGGAGACACAGAAATGGGAGATCTCTCGAATGTAACATCTTACGAGAGTCCCGG 434
 483 CCCCAGG 489
 435 CTCACAG 441

RESULT 8

US-08-494-168-1
 Sequence 1, Application US/08494168

GENERAL INFORMATION:
 PATENT NO. 5731192
 APPLICANT: Reeders, Stephen T.
 APPLICANT: Zhou, Jing
 TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
 TITLE OF INVENTION: of Detecting Collagen Deficiency
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington, D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/494,168
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/112,465
 FILING DATE: 27-AUG-1993
 ATTORNEY/AGENT INFORMATION:

NAME: SAXE, Bernhard D.
 REGISTRATION NUMBER: 28,665
 REFERENCE/DOCKET NUMBER: 40397/104/BAHR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5102 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join(2..82, 86..97, 101..4399, 4403..4420, 4424
 ..4465, 4469..4876, 4880..5101)
 US-08-494-168-1

Query Match 7.3%; Score 100; DB 1; Length 5102;
 Best Local Similarity 52.4%; Pred. No. 5.3e-17;
 Matches 220; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

468 ACAAGTCCACCGGCGCCCGAGGGTCCAGAGGTGACAGAGATCCACAGACCCCTGG 527
 2380 AATGTGCTCCGGGGGAGACAGGCTTACAGAGATTAACAGGGACAAAGATTTCTTG 2439
 528 CCCACTGGCAACAGGACAGAGAGAGAGAGAGGGGAGCTTGACCACTGGCCCTGC 587
 2440 AGACTCTGCTCTCCAGACTCAAGAGGTGTGACAGGGAGAGGCTGTCTAGGCTCA 2499
 588 GGGTGAGAGAGGCCCAATGTGACAGCTGCTCCCGGAGAGCGTGGGGAAGAGATC 647
 2500 AGGTGAGCGGGGAGCGCTTGAGACACAGAGAGAGTGGGACACAGGACCCCAAGATC 2559
 648 TAAAGCTCCCAAGGCGCCCAAGGCTCCGTTGCTCCGAGAAAGCCCGCTCAGG 707
 2560 TAGTGTTCATTTGGATCAAGGGCAATCTGGGCTCCAGAGAGCACAGGCTTCCAGG 2619
 708 CCCCAGTGGGAGCCAGGCGCCCGGCGCCACAGGCAAGAGGACTCCCGGCGCTCA 767
 2620 CATCTCAGAGCATCTCTGAGAAAGAGAGAGGCAAGAGGCTCTCTGATCAAT 2679
 768 GGGCCCTCTGCTTCCAGGACTTCCAGGACCTTCCAGGAGCGTGGGGTCTGTGAC 827
 2680 TGTAAAGAAAGGCTGCGAGGCTTAAAGGCTTCTTGAATTCAGGCTTGTAGGACT 2739
 828 TCGGGAGCTGCGAGGCTTGTGCTGCTTACAGGAGCATAGGCGCCCAAGGCGCCCG 887
 2740 GAAAGGAAAGCCAGGCTCTCCAGGAGGTGCTGCTGGGTTGCCAGGCTCTCTGAC 2799

RESULT 9

US-07-621-091G-1
 Sequence 1, Application US/07621091G

GENERAL INFORMATION:
 PATENT NO. 5424408
 APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
 APPLICANT: G.
 TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
 TITLE OF INVENTION: Polynucleotides
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Yale University, Office of Cooperative Research
 STREET: 246 Church Street
 CITY: New Haven
 STATE: Connecticut
 COUNTRY: U.S.A.
 ZIP: 06510
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
 COMPUTER: Apple Macintosh
 OPERATING SYSTEM: Macintosh OS7.0

```
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/621,091G
FILING DATE: 11/30/90
CLASSIFICATION: 424
PRIOR APPLICATION DATA: NO. 5424408 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Barth, Richard S.
REGISTRATION NUMBER: 28180
REFERENCE/DOCKET NUMBER: 900983/RB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: NO. 5424408 known
CHROMOSOME/SEGMENT: NO. 5424408 known
PUBLICATION INFORMATION: NO. 5424408e
US-07-621-091G-1
```

Query Match 7.3%; Score 99.6; DB 1; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

```
472 GGTCCACCGGGCCCCAGGGTCCCAAGAGTGACAGAGATCCCAAGGCCCTGGGCCCA 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 GGAGCACCAGAGTGTCCCGGTCAAGCCAGGGGCCAGAGTGATCTGTATGTGATTT 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 ACTGGCAACAAGGACAGAAAGAGAGAGAGGGAGCTGGACCACTGGCCCTGCGGT 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 CCAGGCAATGAAGGAGAGAGGGTAATTTCAGGATTTCCAGGACCACTGGACCTCCAGGG 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 GAGAGAGCCCAATTGGACCACTGGTCCCGGGAGAGCGTGGCCGCAAAAGATCTAA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 CAAGTGAGACCAAAAGGACCACTGGAGTACGTGAGAGCGTGGCACAGTGAAGATCATC 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
652 GGTCCACCGGGCCCCAGGGTCCCAAGAGTGACAGAGATCCCAAGGCCCTGGGCCCA 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 TCCCTTCAGAGAACCCAGGCCCACTGTTCAGCTGGAAGACAGGAGATGCAAGGAA 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
712 AGTGGAGACCCAGGCCCCCGGGGCCACAGGCAAAAGAGGACTCCCGGCCCTCAAGGC 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
523 CCGGGGCCCCAGGACCAAGAGATCCAGGACCTGTGGGCCAAAGGTAACCAAGGG 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
772 CTTCTGGCTTCCAGGAGACTTCAGAGGACCGCTTGGGAGCGCTGGGGTGCCTGACCTGG 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 GAGATGTGTCACACAGGAATCTCTGGACCAACTGAGCAAAAGGCAAAAGGTTGTAA 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 GGAAGTCCAGAGCTTGTCTGGGGTACCAAGGATCCAGGCCCAAGGGCCCCCGGCCCT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 GGAAGCAAGGACCACTGATCCGATGGCTCCAGGCTTGAAGGGGAAACCTGGAGAC 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
892 CTTGGCCCATCAGAGCGGTGG 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
703 ACTGGACCACTGCAGCAGGGG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
US-08-399-889-1
; Sequence 1, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-08-399-889-1
```

Query Match 7.3%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

```
472 GGTCCACCGGGCCCCAGGGTCCCAAGAGTGACAGAGATCCCAAGGCCCTGGGCCCA 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 GGAGCACCAGAGTGTCCCGGTCAAGCCAGGGGCCAGAGTGATCTGTATGTGATTT 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 ACTGGCAACAAGGACAGAAAGAGAGAGAGGGAGCTGGACCACTGGCCCTGCGGT 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 CCAGGCAATGAAGGAGAGGGTAATTTCAGGATTTCCAGGACCACTGGACCTCCAGGG 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 GAGAGAGCCCAATTGGACCACTGGTCCCGGGAGAGCGTGGCCGCAAAAGATCTAA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 CAAGTGAGACCAAAAGGACCACTGGAGTACGTGAGAGCGTGGCACAGTGAAGATCATC 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
652 GGTCCACCGGGCCCCAGGGTCCCAAGAGTGACAGAGATCCCAAGGCCCTGGGCCCA 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 TCCCTTCAGAGAACCCAGGCCCACTGTTCAGCTGGAAGACAGGAGATGCAAGGAA 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
712 AGTGGAGACCCAGGCCCCCGGGGCCACAGGCAAAAGAGGACTCCCGGCCCTCAAGGC 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
523 CCGGGGCCCCAGGACCAAGAGATCCAGGACCTGTGGGCCAAAGGTAACCAAGGG 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
772 CTTCTGGCTTCCAGGAGACTTCAGAGGACCGCTTGGGAGCGCTGGGGTGCCTGACCTGG 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 GAGATGTGTCACACAGGAATCTCTGGACCAACTGAGCAAAAGGCAAAAGGTTGTAA 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 GGAAGTCCAGAGCTTGTCTGGGGTACCAAGGATCCAGGCCCAAGGGCCCCCGGCCCT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 GGAAGCAAGGACCACTGATCCGATGGCTCCAGGCTTGAAGGGGAAACCTGGAGAC 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
892 CTTGGCCCATCAGAGCGGTGG 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
703 ACTGGACCACTGCAGCAGGGG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11
US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B

;; CURRENT APPLICATION NUMBER: US/09/167,364
;; CURRENT FILING DATE: 1998-10-07
;; EARLIER APPLICATION NUMBER: 08/399889
;; EARLIER FILING DATE: 1995-03-07
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Calf
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-167-364-1

Query Match 7.3%; Score 99.6; DB 3; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 472 GGTCCACCGGCCCCAGGGGTCACAGAGTGACAGAGATCCAGGGACCCCTGGCCCA 531
DB 283 GGAGCACAGGTGTCCCGGTACAGAGGGCCAGAGTGATCTGGATTCTATGATTT 342
QY 532 ACTGGCAACAGGACAGAAAGAGAAAGGGGAGCTGAGACACCTGGCCCTGCGGGT 591
DB 343 CCAGGATGAAAGGAAAGGTAATTCAGATTTCCAGAGACACCTGGACCTCCAGGG 402
QY 592 GAGAGAGGCCCAATTGACACAGCTGTCCCGGAGAGAGCTGGCGGCAAGATCTAA 651
DB 403 CAAGTGAGCAAAAGGACACCTGGATGACGTGAGAGACCTGGCAGTGAAGATC 462
QY 652 GGCTCCACAGGCCCCAAAGGCTCCCGTGGTCCCTGGGAAAGCCCGCCCTGAGGGCCC 711
DB 463 TCCCTTCAGAGAAAGCCAGGCCACCTGGTTCAGCTGAGAAACAGGATGACAGAA 522
QY 712 AGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGACTGCCCGCTCAGGGC 771
DB 523 CCGGGGCCCCAGAGACACAGAGATCCAGACCTGTGGGCAAAAGGTAACAGGG 582
QY 772 CCTCTGGCTTCAGGAGACTTCAGGACACCTGGGGAGCCTGGGCTGACCTCG 831
DB 583 GAGGATGTTCACACAGAACTCCTGGACCACTGGAGAAAGCAAAAGTTGTAA 642
QY 832 GGAATGCGGCTTGGCTGGGTACAGGATGCGAGGCCAGAGGCCCGCCGCGCT 891
DB 643 GGAGAGCAAGAGACACCTGGATCCGATGGCTTCAGGCTTGAAGGGAAACCTGAGAC 702
QY 892 CCTGGCCATCAGAGCGGTGG 913
DB 703 ACTGACACACTGACAGCAGGG 724

RESULT 12
US-09-439-897-1

;; Sequence 1, Application US/09439897
;; Patent No. 6277558
;; GENERAL INFORMATION:
;; APPLICANT: Hudson, Billy G
;; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
;; FILE REFERENCE: 95-1263-C
;; CURRENT APPLICATION NUMBER: US/09/439,897
;; CURRENT FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 65
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-439-897-1

Query Match 7.3%; Score 99.6; DB 4; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 472 GGTCCACCGGCCCCAGGGGTCACAGAGTGACAGAGATCCAGGGACCCCTGGCCCA 531
DB 283 GGAGCACAGGTGTCCCGGTACAGAGGGCCAGAGTGATCTGGATTCTATGATTT 342
QY 532 ACTGGCAACAGGACAGAAAGAGAAAGGGGAGCCTGAGACACCTGGCCCTGCGGGT 591
DB 343 CCAGGATGAAAGGAAAGGTAATTCAGATTTCCAGAGACACCTGGACCTCCAGGG 402
QY 592 GAGAGAGGCCCAATTGACACAGCTGTCCCGGAGAGAGCTGGCGGCAAAAGATCTAA 651
DB 403 CAAGTGAGCAAAAGGACACCTGGATGACGTGAGAGACCTGGCAGTGAAGATC 462
QY 652 GGCTCCACAGGCCCCAAAGGCTCCCGTGGTCCCTGGGAAAGCCCGCCCTCAGGGCCC 711
DB 463 TCCCTTCAGAGAAAGCCAGGCCACCTGGTTCAGCTGAGAAACAGAGATGCAAGGAA 522
QY 712 AGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGACTCCCGGCTCAGGGC 771
DB 523 CCGGGGCCCCAGAGACACAGGATCCAGGACCTGTGGGCAAAAGGTAACAGGG 582
QY 772 CCTCTGGCTTCAGGAGACTTCAGGACACCTGGGGAGCCTGGGCTGACCTCG 831
DB 583 GAGGATGTTCACACAGAACTCCTGGACCACTGGAGAAAGCAAAAGTTGTAA 642
QY 832 GGAATGCGGCTTGGCTGGGTACAGGATGCGAGGCCAGAGGCCCGCCGCGCT 891
DB 643 GGAGAGCAAGAGACACCTGGATCCGATGGCTTCAGGCTTGAAGGGAAACCTGAGAC 702
QY 892 CCTGGCCATCAGAGCGGTGG 913
DB 703 ACTGACACACTGACAGCAGGG 724

RESULT 13
US-08-392-367B-1

;; Sequence 1, Application US/08392367B
;; Patent No. 5691197
;; GENERAL INFORMATION:
;; APPLICANT: Trygveason, Karl
;; APPLICANT: Eliomaa, Outi
;; APPLICANT: Kangas, Maarit
;; TITLE OF INVENTION: An Insolated DNA Sequence For a
;; Patent No. 5691197
;; TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
;; TITLE OF INVENTION: a Collagenous Domain and the
;; TITLE OF INVENTION: Polypeptide Chain Encoded by
;; TITLE OF INVENTION: such a Sequence
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
;; ADDRESSEE: Minnich & McKee
;; STREET: 1100 Superior Avenue
;; STREET: Suite 700
;; CITY: Cleveland
;; STATE: Ohio
;; COUNTRY: U.S.A.
;; ZIP: 44114-2518
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch,
;; MEDIUM TYPE: 720 Kb storable
;; COMPUTER: IBM PS/2, Model 35 SX
;; OPERATING SYSTEM: DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/392,367B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Minnich, Richard J.


```

Sequence 11, Application US/08555669
Patent No. 5773248

GENERAL INFORMATION:
APPLICANT: Brewton, Richard G.
APPlicant: Mayne, Richard
TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue Of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/555,669
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2543 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 47..2098
US-08-555-669-11

```

Query Match	7.3%	Score 99.4	DB 1	Length 2543
Best Local Similarity	51.5%	Pred. No.5.6e-17		
Matches 229	Conservative	0	Mismatches 216	Indels 0
			Gaps	0
QY	CAAGTCCACCGGGGCCCCAGGGGATCCAAAGATGTGACAGAGATGCCAGGAGACCCCTGGC	528		
Db	CCAGTTCGCCCAAGGGGCCCCCTGGAAATGCCAGGGTTCMAAGGACCCACTGTATCAAAAGC	643		
QY	CCAACTGGCAACMAAGGACAGAAAGAGAGAAAGGGGAGCCCTGGACCACTGGCCCTGGC	588		
Db	GAGCAGGGGGAAGTGCAGCAAGGACGGCGAAGAGGGTGACCTTGGCCCCCTGGGCCCGCC	703		
QY	GGTAGAGAAAGGCCCAATTGGACCACTGTGTCCCCCGGAGAGCCTGTGCGCGCAAAAGATT	648		
Db	GGCCTCCCGCGACAGGTGGGGCTGCAAGGGCCCCCGGGGATTTCAGAGATGCCAGGGCCA	763		
QY	AAAGGCTCCCAAGGGGCCCCAAAGGCTCCCGGTGTCCCTGGGAGAGGCGGGCCCTCGAGGC	708		
Db	CTCGGGCCCCCTGGGGAGACCGGGGTCCCAATTGGGTTTCCAGGGCCGCTTGATTCAGGA	823		
QY	CCCACTGGGGGACCCAGGCCCCCGGGGCCACACAGGCAAAAGAGGACTCCCGGACCTTCAG	768		
Db	GCGCCTGGGAAAGCGGGTGACCGAAGGCGAGAGGGGCCCGAAGAGGTTCTCCGGCCCCCAG	883		
QY	GACCTCTCTGCTTCCAGGAGACTCAGGGACCGTTGGGGAGCCTGGGGTGTCTGTGACT	828		
Db	GGTGACTCTGCGAGACTGTGTCCAAAGGAAACCCCGGAGTGGCGCGGCCAAGGGAGAG	943		
QY	CGGGACTGCCAGGCTTGGGCTTACCAAGCATATCCAGGCCCCCAAGGGCCCCCCCCGC	888		

Db 944 CCGGGCATGCGCAGGCAAGACGGCCAGAAATGGCTGTCCAGGACATGATGGCCAGAAGGA 1003

QY 889 CCTCTGAGCCCATCAGAGAGCGGTGG 913

Db 1004 GAGGCTGTGTGCAACGAGTGTCTCGG 1028

```
Search completed: March 21, 2003, 08:25:41
Job time : 61.326 secs
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 208.91 Seconds
(without alignments)
5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_325_1695
Perfect score: 1371
Sequence: 1 atgaacagccagctcaactc.....cagtaactgtcatcgtcatla 1371

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCUTS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1369.4	99.9	2930	10	US-09-745-763-198
2	111	8.1	6728	10	US-09-954-456-782
3	111	8.1	6728	10	US-09-880-107-3946
4	110.6	8.1	1133	9	US-09-924-340-57
5	110.6	8.1	1133	9	US-09-992-600A-57
6	109.4	8.0	2691	10	US-09-925-302-64
7	107.2	7.8	2192	10	US-09-925-301-42
8	107	7.8	5086	10	US-09-880-107-7947
9	107	7.8	5145	9	US-09-925-299-206
10	107	7.8	5145	10	US-09-925-299-206
11	107	7.8	5432	12	US-10-044-090-22
12	106.8	7.8	4908	9	US-10-001-887-33
13	106	7.7	2542	9	US-09-954-531-961
14	106	7.7	2542	10	US-09-964-824A-255
15	105.4	7.7	5416	10	US-09-954-456-786
16	105.4	7.7	5416	10	US-09-880-107-2094
17	105.2	7.7	6158	10	US-09-919-497-6
18	105.2	7.7	6158	10	US-09-954-456-762
19	104.2	7.6	4149	12	US-10-044-090-104

20	102.4	7.5	1486	10	US-09-925-302-247	Sequence 247, App
21	101.6	7.4	3226	10	US-09-954-456-725	Sequence 725, App
22	101.4	7.4	3690	12	US-10-044-090-448	Sequence 448, App
23	98.6	7.2	2158	9	US-10-001-835-98	Sequence 98, App
24	97.8	7.1	2520	10	US-09-880-107-3685	Sequence 3685, App
25	95.8	7.0	1485	10	US-09-925-302-246	Sequence 246, App
26	95.2	6.9	1797	9	US-09-978-295A-613	Sequence 613, App
27	95.2	6.9	1797	9	US-09-978-697-613	Sequence 613, App
28	95.2	6.9	1797	9	US-09-978-192A-613	Sequence 613, App
29	95.2	6.9	1797	9	US-09-998-832A-613	Sequence 613, App
30	95.2	6.9	1797	9	US-09-978-189-613	Sequence 613, App
31	95.2	6.9	1797	9	US-10-174-590-331	Sequence 331, App
32	95.2	6.9	1797	9	US-10-176-758-331	Sequence 331, App
33	95.2	6.9	1797	9	US-10-175-737-331	Sequence 331, App
34	95.2	6.9	1797	9	US-10-173-706-331	Sequence 331, App
35	95.2	6.9	1797	9	US-10-175-728-331	Sequence 331, App
36	95.2	6.9	1797	9	US-10-175-752-331	Sequence 331, App
37	95.2	6.9	1797	9	US-10-176-482-331	Sequence 331, App
38	95.2	6.9	1797	9	US-10-176-757-331	Sequence 331, App
39	95.2	6.9	1797	9	US-10-176-913-331	Sequence 331, App
40	95.2	6.9	1797	9	US-10-180-552-331	Sequence 331, App
41	95.2	6.9	1797	9	US-10-180-557-331	Sequence 331, App
42	95.2	6.9	1797	9	US-10-173-700-331	Sequence 331, App
43	95.2	6.9	1797	9	US-10-174-572-331	Sequence 331, App
44	95.2	6.9	1797	9	US-10-174-579-331	Sequence 331, App
45	95.2	6.9	1797	9	US-10-174-582-331	Sequence 331, App

ALIGNMENTS

RESULT 1
US-09-745-763-198
Sequence 198, Application US/09745763
Patent No. US2002006394A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
LaValle, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merbery, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 219
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 198:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match 99.9%; Score 1369.4; DB 10; Length 2930;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGAACAGCCAGCTCACTATTTCACAGGTCAATGAGAAACATCACCATTATCTCCAA 60
DB 921 ATGAACAGCCAGCTCACTATTTCACAGGTCAATGAGAAACATCACCATTATCTCCAA 980
QY 61 GCCAACAGCAGAGAACCTGGAAGACCTCAGACCTTACACAAAGATGAGAGAAATAGACA 120
DB 981 GCCAACAGCAGAGAACCTGGAAGACCTCAGACCTTACACAAAGATGAGAGAAATAGACA 1040
QY 121 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTTTTGAGACGATATTGTGAC 180
DB 1041 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTTTTGAGACGATATTGTGAC 1100
QY 181 ATCATTTGACATATCACTTACACAGCCCACTGCGGACGTGACACATCTTAAT 240
DB 1101 ATCATTTGACATATCACTTACACAGCCCACTGCGGACGTGACACATCTTAAT 1160
QY 241 GAAGTCAGACCACTTGCACAGATACCTTACCAAAACACAGATGATGACCTCCTTG 300
DB 1161 GAAGTCAGACCACTTGCACAGATACCTTACCAAAACACAGATGATGACCTCCTTG 1220
QY 301 AATTAATACCTTGCCCAACATCCGTTGGATTCTGTTTCTCTCAGAGATGACAAAGATTG 360
DB 1221 AATTAATACCTTGCCCAACATCCGTTGGATTCTGTTTCTCTCAGAGATGACAAAGATTG 1280
QY 361 ATGAGGTGAGGTAGACATGAAGTGAAGTCACTTATCATGATTTATGAGAAATAG 420
DB 1281 ATGAGGTGAGGTAGACATGAAGTGAAGTCACTTATCATGATTTATGAGAAATAG 1340
QY 421 CTAGTAGACTCAAGCATGCTGAGCTCATCAAGATTTTCAATACTACAAAGTCCACCG 480
DB 1341 CTAGTAGACTCAAGCATGCTGAGCTCATCAAGATTTTCAATACTACAAAGTCCACCG 1400
QY 481 GGGCCCAAGGGGTCCAGAGGTGACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 540
DB 1401 GGGCCCAAGGGGTCCAGAGGTGACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 1460
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGGACCTGAGACACCTGGCCCTGGGGTGAGAGAGG 600
DB 1461 AAGGACAGAAAGAGAGAGAGGGGAGGACCTGAGACACCTGGCCCTGGGGTGAGAGAGG 1520
QY 601 CCAATTGGACACGTGTCTCCCGGAGAGCTGGGCGGCAAAAGATCTAAAGGCTCCAG 660
DB 1521 CCAATTGGACACGTGTCTCCCGGAGAGCTGGGCGGCAAAAGATCTAAAGGCTCCAG 1580
QY 661 GGGCCCAAGGCTCCGTTGTTCCCTGGGAGAGCCGCGCTTACAGGCCCCAGTGGGGAC 720
DB 1581 GGGCCCAAGGCTCCGTTGTTCCCTGGGAGAGCCGCGCTTACAGGCCCCAGTGGGGAC 1640
QY 721 CCAAGGCCCCCGGGGCGGACAGAGAAAGAGAGGATCCCGGCGCTCAAGGCCCCCTCCG 780
DB 1641 CCAAGGCCCCCGGGGCGGACAGAGAAAGAGAGGATCCCGGCGCTCAAGGCCCCCTCCG 1700
QY 781 TTCACAGGACTTACAGGACACCTGTTGGGAGCTTGGGGTGCCTGAGACCTCGGGGACTGCCA 840
DB 1701 TTCACAGGACTTACAGGACACCTGTTGGGAGCTTGGGGTGCCTGAGACCTCGGGGACTGCCA 1760
QY 841 GGGCTTGGCTGGGTACAGGACATGCAAGGCCCAAGGGCCCCCGGCGCTCTCTGGGCCCA 900
DB 1761 GGGCTTGGCTGGGTACAGGACATGCAAGGCCCAAGGGCCCCCGGCGCTCTCTGGGCCCA 1820
QY 901 TCAGAGACGGTGGTCCCGCTGGGCTGCAAGATGAGCAACCCCGGACCCGAGAGACAAAT 960
DB 1821 TCAGAGACGGTGGTCCCGCTGGGCTGCAAGATGAGCAACCCCGGACCCGAGAGACAAAT 1880
```

```
QY 961 GGCTGCCCGCTCCTCTGGAAGACTTACAGACAAATGCTACTATTTTCAGTTGAGAAA 1020
DB 1881 AGCTGCCCGCTCCTCTGGAAGACTTACAGACAAATGCTACTATTTTCAGTTGAGAAA 1940
QY 1021 GAAATTTTGGAGATCAAGCTTTTCTGTGGAAGACAGTCTTCACATCTTGTTCATA 1080
DB 1941 GAAATTTTGGAGATCAAGCTTTTCTGTGGAAGACAGTCTTCACATCTTGTTCATA 2000
QY 1081 AACACTAGAGAGGAGACAGCAATGATATAAAACACATGTTAGGAGACAGACCCCTGG 1140
DB 2001 AACACTAGAGAGGAGACAGCAATGATATAAAACACATGTTAGGAGAGAGACCTGG 2060
QY 1141 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAAATGGAAGTGGCTGATGGACATCTCA 1200
DB 2061 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAAATGGAAGTGGCTGATGGACATCTCA 2120
QY 1201 GACTACAAAATTTGAAAGCTGAGACAGCGGATTAAGTGGGCTATGCGCATGGGCCAGGA 1260
DB 2121 GACTACAAAATTTGAAAGCTGAGACAGCGGATTAAGTGGGCTATGCGCATGGGCCAGGA 2180
QY 1261 GAAGACTGTGCTGGTTGATTATGCTGGGCGAGTGAAGCATTTCCATGTGAAGACGTC 1320
DB 2181 GAAGACTGTGCTGGTTGATTATGCTGGGCGAGTGAAGCATTTCCATGTGAAGACGTC 2240
QY 1321 AATACTTCATTTGCGAAAAGACAGGAGACAGTACTGTCATCTGCATTA 1371
DB 2241 AATACTTCATTTGCGAAAAGACAGGAGACAGTACTGTCATCTGCATTA 2291
```

RESULT 2

US-09-954-456-782
Sequence 782, Application US/09954456
Patent No. US20020115057A1

GENERAL INFORMATION:

APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
CURRENT FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 782
LENGTH: 6728
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-782

Query Match

Best Local Similarity 8.1%; Score 111; DB 10; Length 6728;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

[illegible]

```

RESULT 3
US-09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3946
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z74615
US-09-880-107-3946

Query Match      8.1%; Score 111; DB 10; Length 6728;
Best Local Similarity 53.0%; Pred. No. 2,6e-20;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0.

Oy 467 TACAGAGTCCACCGGCGCCAGGCGTCCAAAGAGGTGACAGAGATCCCGAGGAGCCCCCTG 526
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 721 TCCAAAGGTCCCTTGCTGTGAGCTTGCGAGAGCTTGAGAGCTTCAGATGGCTCCCGGAG 780

Oy 527 GCCCACTGCGCAACAAGGAGCAAGAAAGAGAGAAGGGGGAGCGCTGAGACCACTGGCCCTG 586
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 781 GTCCCCCAGGTCCCTTGGAAGAAGATGAGATGATGGGGAGACCTGGAAGAACCCTGGTCTG 840

Oy 587 CGGGTGAGAGAGGCCCAATTGGAGCAGCTGTGTCCTCCCGGAGAGAGCGCTGGCGCAAAAGAT 646

```

Db	841	CTGTGAGCCTGGGCTCCTCGGSCCTCAGGAGTGTCTGAGAGATTGGCCGGAAACACTGGCC	11
QY	647	CTAAGGCTCCCAAGGGCCCCAAAGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGG	706
Db	901	TCCCTGGAATGAAGAGGACACAGAGCTTTTCAGTGGTTGGATGTGTCACAAAGGAATGCTG	960
QY	707	GCCCCATGGGGAGACCAGGGCCCCCGGGGGCCACACAGGCACAAAGGACATCCCGGGCCCTC	766
Db	961	GTCCTGCTGTGCTCTAAGGGAGAGCTGGAGCCCTGTGTGTAATATGAGACTCTGTGTGAGA	1020
QY	767	AGGGCCCTCTCGGCTTCCAGGAGACTTCAGGGGACACCGTTTGGGAGGCTGGGGTGCCTGAC	826
Db	1021	TGGGCCCCCGTGGGCTTCGCTGAGAGAGAGTGCCTCGAGACCCCTCGCCCTGTGGTG	1080
QY	827	CTGGGGGACATGCCAGGCTTGCCTTGGGGGTACACAGGATGTCAGAGGCCCAAGGGCCCCCCC	886
Db	1081	CTCCTGGAAATGATGTGTCTACGTGAGTGCCGCGGGGCCCTCGTCCACACGGCCCCCCTG	1140
QY	887	GCCCTCTGCGCCATCAGAGAGCGGTG	913
Db	1141	GTCCTCTGGCTTCCCTGGTGTGCTGG	1167

```

RESULT 4
US-09-924-340-57
; Sequence 57, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.Reg
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
; US-09-924-340-57

Query Match      8.1%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred.No.1.2e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 121

OY 435 GCATGCTACGCTCATCAAGATTMTTACATCTACTACAGAGTCCACCGGCGCCCGAGGGGTCC 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 109 GCAAAGACGAGGCGCCCGCCAGGATCTGAAGGCTCCAGGCCCGCCCGAGCGCCCGAGCGGTCC 168

OY 495 AAGAGGTACACAGAGATCCCGACGAGACCCCGCTGGCCCACTGGCCACACAAAGGACAAAGG 554
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 169 CAGAGGACGAGCGAGGACCCCAAGGTAACTCCGGTGAGAAAGGCGACGACCAAGATTTCAAAG 228

```


OY 555 AGAAGAGGGAGCTGGACCTGGCCCTGGCGGTGAGAGAGCCCAATTGGACG 614
DB 229 CCAAGCCAGGCTTTCGGGCGCCACCGGGTCCCTGGATTCACCAAGTTGGATCACC 288
OY 615 TGTGCCCCCGG-----AGAGCTGGCGGCAAGGATCTAAAGGCTCCAGGGCCC 665
DB 289 TGGCCCACTGGCCCTCAAGCAGAGAGGAGGCAAGGATTCAGAGCCATCAGGCT 348
OY 666 CAAAGGCTCCGTTGTTCCCTGGAGAGCCGCGCTCAGAGGCCCGAGTGGAGCCAGG 725
DB 349 GCCTGGTCCCTTGGGCGCCACCGGACCTCTGGGATTCAGAGGCCCGCTGATGG 408
OY 726 CCCCCGGGCCCAAGCAGGCAAGAGGAGCTCCCGGCTCAGAGGCCCTGCTTCA 785
DB 409 TTTGATGAGGAGAGGATGCAAGCTGCTGAGGGGGGAGCCCTGCTGCTGAGCCCC 468
OY 786 GGGACTCAGGAGCCTTTGGGAGGCTGGGCTGACCTGGGAGCTGGAGGCT 845
DB 469 TGGACTCATGGGACACCGGCTTTAAGGGGAAACAGACATCTGCTCCAGGACC 528
OY 846 GCTGAGGCTACCAAGCATGTCAGAGGCCCGCGGCTCTGAGGCTGAGG 905
DB 529 TAAGGCTGACTGTGCAACACGATCTCTGCGAGCAGTGGCGGCTGGCGAGAGG 588
OY 906 AGCGGTGTGCTCCCTGGCCCTGCAG 930
DB 589 TGAACCTGTGTGCTGAGGAGCCCGCAG 613

RESULT 5
US-09-992-600A-57
Sequence 57, Application US/09992600A
Publication No. US2003027161A1
GENERAL INFORMATION:

APPLICANT: Benjamin, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNAs AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.054.DIV
CURRENT APPLICATION NUMBER: US/09/992,600A
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 114
SOFTWARE: JPatent
SEQ ID NO 57

LENGTH: 1133
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..475
FEATURE:
NAME/KEY: CDS
LOCATION: 476..964
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 965..1133
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1101..1106
FEATURE:
NAME/KEY: polyA_site

LOCATION: 1118..1133
US-09-992-600A-57

Query Match 8.1%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 1,2e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

OY 435 GCATGTGACCTCATCAAGATTTTACAAATACAAAGTCAACGAGGCCCGAGGGGTCC 494
DB 109 CAAAGACAGAGGCGCCCGCCAGATCTGAAAGGCTCCAGGCGCCCGCCAGCGGGTCC 168
OY 495 AAGAGTGAACAGAGATCCAGAGGACCCCTGGCCCACTGGCAACAGGAGACAGAAAG 554
DB 169 CAGAGGAGACAGAGAGACCCCAAGGTAACTCGGTGAGAAAGGGAGCCAGGATTTCAAG 228
OY 555 AGAAGAGGGAGGAGCTGAGACCTGGGCGGTGAGAGAGGCCCAATTGGAGCAG 614
DB 229 CCAAGCCAGGCTTTCGGGCGCCACCGGGTCCCTGGATTCACCAAGTTGGATCACC 288
OY 615 TGTGCCCCCGG-----AGAGCTGGCGGCAAGGATCTAAAGGCTCCAGGGCCC 665
DB 289 TGGCCCACTGGCCCTCAAGCAGAGAGGAGGCAAGGATTCAGAGGCCCATCAGGCT 348
OY 666 CAAAGGCTCCGTTGTTCCCTGGAGAGCCCGGCTCAGAGGCCCGAGTGGAGCCAGG 725
DB 349 GCCTGGTCCCTTGGGCGCCACCGGACCTCTGGGATTCAGAGGCCCGCTGATGG 408
OY 726 CCCCCGGGCCCAAGCAGGCAAGAGGAGCTCCCGGCTCAGAGGCCCTGCTTCA 785
DB 409 TTTGATGAGGAGAGATGCAAGCTGCTGAGGGGGGAGCCCTGCTGCTGAGCCCC 468
OY 786 GGGACTCAGGAGCCTTTGGGAGGCTGGGCTGACCTGGGAGCTGGAGGCT 845
DB 469 TGGACTCATGGGACACCGGCTTTAAGGGGAAACAGACATCTGCTCCAGGACC 528
OY 846 GCTGAGGCTACCAAGCATGTCAGAGGCCCGCGGCTCTGAGGCTGAGG 905
DB 529 TAAGGCTGACTGTGCAACACGATCTCTGCGAGCAGTGGCGGCTGGCGAGAGG 588
OY 906 AGCGGTGTGCTCCCTGGCCCTGCAG 930
DB 589 TGAACCTGTGTGCTGAGGAGCCCGCAG 613

RESULT 6
US-09-925-302-64
Sequence 64, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 64

LENGTH: 2691
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (2653)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (2667)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (2683)

OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-64

Query Match 8.0%; Score 109.4; DB 10; Length 2691;
Best Local Similarity 54.4%; Pred. No. 4,3e-20;
Matches 240; Conservative 1; Mismatches 197; Indels 3; Gaps 1;

QY 469 CAGGTCCACCGGCCCCCAGAGGCTCCAGAGGATCCAGAGGACCCCTGGC 528
DB 31 CCAGGAGACCCGCTCCAGAGGACCCCTGGATCCAGAGGACCCCTGGT 90
QY 529 CCACTGGCAACAGAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
DB 91 ATTGGCTCTCCAGAAATACAGGCTCTAAGGCTGATATGGGGCTCCAGAGATTC 150
QY 589 GGTGAGAGAGGCGCAATTTGGACCACTGGTCCCCCGAGAGAGGCTGGCGCAAGATCT 648
DB 151 GGATTTCAAGGTCCAAAGAGGCTTCTGGCTCCAGAGGAAATTAAGGTATCAAGGCGAT 210
QY 649 AAGGCTCCAGAGGCCCCCAAGGCTCCGCTGGTCCCTGGAGAGAGAGAGAGAGAG-- 706
DB 211 CAGGCTCCCGGAGAGCTTAAGGTCTCCGCTCTCTGGCCCCCAGAGCTCTTACGAC 270
QY 707 -GCCCAAGTGGGAGCCAGGCCCCCGGAGCCACAGGCAAGAGAGAGAGAGAGAGAGCT 765
DB 271 ATCATCAAGAGGAGAGCCCGGCTCCCTGGCTCGAGAGGCCCCCAGAGGCTGAAGGCTT 330
QY 766 CAGGCCCCCTCTGGCTTCCAGAGGACTTCAAGGACCGCTTGGGAGAGCTGGGGCTCTGA 825
DB 331 CAGGAGACTCCAGAGGCCCCAGAGAGGCTTACAGGATGGTGGGTATACCTGGA 390
QY 826 CCGGAGGAGTCCAGAGGCTTCCCTGGGATACAGGATGCGAGGCCCCCAAGGAGAGAGAG 885
DB 391 CCGCAGGATCTCTGGGCTTACGCTGCCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 450
QY 886 GGCCCTCTGGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
DB 451 GGCCCTCTGGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471

RESULT 7

US-09-925-301-42
; Sequence 42, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 42
; LENGTH: 2192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-42

Query Match 7.8%; Score 107.2; DB 10; Length 2192;
Best Local Similarity 53.5%; Pred. No. 1,6e-19;
Matches 214; Conservative 0; Mismatches 228; Indels 10; Gaps 2;

QY 472 GGTCCACCGGCCCCCAGGAGTCCAGAGAGATCCAGAGGAGAGAGAGAGAGAGAGAGAG 531
DB 638 GGGCCCGGTGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 697
QY 532 ACTGGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 591
DB 698 GGTGCAAGAGTGTGATGCTGTGAGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 757

QY 592 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
DB 758 CCGTGTGGCCCGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
QY 652 GGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 711
DB 818 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 712 AGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771
DB 877 -----CCCTGGCCCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTGCT 928
QY 772 CCGTCTGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 831
DB 929 CCGTCTGGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 988
QY 832 GGAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
DB 989 GGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1048
QY 892 CCGTGGCCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 950
DB 1049 GTTGTGCTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1108
QY 951 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 982
DB 1109 GACTTACAGCTTCTGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140

RESULT 8

US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1) (5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Query Match 7.8%; Score 107; DB 10; Length 5086;
Best Local Similarity 51.3%; Pred. No. 2,9e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 465 ACTCAAGTCCACCGGCCCCCAGGAGTCCAGAGAGTGAAGAGATCCAGAGAGAGAGAGAGAG 524
DB 709 ACAGCCCGGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 768
QY 525 TGGCCCACTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 584
DB 769 AGGTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 828
QY 585 TGGGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644

D b	829	AGCTGGTGCCTGCGGACGTAGTGAAATGAGGATGCCCGGTCCGTGCCTGACTGTCGCCATTGG	888
O y	645	ATCTAAGGCTCCCAAGGGCCCCCAAAGCTCCCGTGTTCCTCGGAAGCCCGGCCCTCA	704
D b	889	GTCCTGTGGCCCTCCACAGCTTCCACAGTGCCTCTGGCCCAAGGTGAAATTTGGAGCTGT	948
O y	705	GGGCCCCAGTGGGGACCCAGAGCCCCCGGGGCCACCAGGCAAGAAGCATCCTCCGGGCC	764
D b	949	TGGTAACGCTGGTCTCTGTGTCCCGCCGGTCCCGCTGGTGAAGTGGGTCTCCAGGCT	1008
O y	765	TCAAGGCCCCTCTGCTTCCAGGACTTCAGGGCACCCGTGGGGGAACCTCGGGGTCTGG	824
D b	1009	CYCGGGCCCGTTTGGACCTCTGTTATCTCTTGAGCAAAAGGCTTACGTGGCCCAAGG	1068
O y	825	ACCTCGGGAGACTGTCAGGCTTGGCTGGGGACAGGCATGCACAGGCCCCCAAGAGGCCCTC	884
D b	1069	TGCTGTGCTGTTCCTCCGGGCTTGTCTGGGGCTCCCGGCTCTCTTGGAAATCCCGGATTTCC	1128
O y	885	CGGCGCTCTCTGGCCCATCAGAGAGCGGTGTGCCCTCTGGCCCTGCAGAAATGAGCCAAACCC	944
D b	1129	TGGCCCTGTTGGTCTCCCGGTCTACTGTTGCCAAGAGACTTTGTGGAGACCTGCTGCC	1188
O y	945	GGC 947	
D b	1189	AGC 1191	

```

RESULT 9
US-09-925-299-206
: Sequence 206, Application US/09925299
: Publication No. US20030040617A9
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA102
: CURRENT APPLICATION NUMBER: US/09/925,299
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05883
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 206
: LENGTH: 5145
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (4)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (17)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (5126)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (5143)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

```

Query Match	7.8%;	Score 107;	DB 9;	Length 5145;
Best Local Similarity	51.3%;	Pred. No. 2.9e-19;		
Matches 248;	Conservative	0;	Mismatches 235;	Indels 0;
			Gaps	0;

QY	ACTCAAGGTCACCGGGGCCCAAGGTCGACGAGAGATCCAGGAGACCCC	52
465	ACTCAAGGTCACCGGGGCCCAAGGTCGACGAGAGATCCAGGAGACCCC <td>52</td>	52
QY	ACAGCCCGCTCCTCGTGTGAAGGGTGAACTCGTGCCTCGTGAATAATGGAACTCC <th>77</th>	77
Db	ACAGCCCGCTCCTCGTGTGAAGGGTGAACTCGTGCCTCGTGAATAATGGAACTCC <td>77</td>	77
712	ACAGCCCGCTCCTCGTGTGAAGGGTGAACTCGTGCCTCGTGAATAATGGAACTCC <td>77</td>	77
QY	TGGCCCAACTGCGCAACAGGACAGAAAGGAGAGAGGGGAGCCTGGACACACTGGCCC <th>584</th>	584
525	TGGCCCAACTGCGCAACAGGACAGAAAGGAGAGAGGGGAGCCTGGACACACTGGCCC <td>584</td>	584

Db	772	AGGTCAAAACAGAGACCCGTGGGCTTCCTGGTGAAGAGAGACGTTGGTGCCCTGGGCC	831
QY	585	TGCGGGTGAAGAGAGGCCCAATTGGTACACGCTGTCTCCCCCGAGAGACGTGGCGGCAAGG	644
Db	832	AGCTGTGTCCCGCTGGCTGAGTATGAATGTGGGTCTCCGTGGGGCTCTGCTGTCCATTGG	891
QY	645	ATCTAAAGAGCTCCCGAGGAGCCCAAGAGCTCCGTGGTTCCCTGGGAAACCCGGGCGCTCA	704
Db	892	GTCGTGTGGGCTCTCAAGGCTTCCCAAGTGTGCCCTGGGCCCAAGAGGTGAATTTGGARCTGT	951
QY	705	GGGCCCCAGTGGGGAGCCAGGCCCCCGGGGCCACCAAGGCAAGAGGACCTCCCGGCC	764
Db	952	TGGTAACGCTGTGTCTCTGTCTGTCCGCGCGGTCTCCCTGGTGAAGTGGGTCTTCCAGGCC	1011
QY	765	TCAGGGCCCTCTGTGCTTCCAGGAGACTCAAGGACACGTTGGGGAGCGGGGTCTCCGG	824
Db	1012	CTCCGGCCCCGTTGGACCTCTGTGTAACTTGAGCAAAAGGCGCTTACTGTGGTCCAGAGG	1071
QY	825	ACCTTCGGGAGCTGTGCATGTGCCTGGGGTACCAAGGACATGCACAGGCCCAAGGSCCCCC	884
Db	1072	TGCTGTGTGGCCCTTCCGGGCGTGTGGGGCTCCCGGCGTCCGTGAACCCCGGGGATATTC	1131
QY	885	CGGCGCTCTGTGGCCATTCAGAGAGCGGTGGTGCCTTCGSCCTTCGCAATGAACCAACCC	944
Db	1132	TGGCGCTGTGTGTGCTCCGCTGCTACTGTGTCCAGAGGACTTGTGGTGAACCTGTGTCC	1191
QY	945	GAC 947	
Db	1192	AGC 1194	

```

1 RESULT 10
2 US-09-925-299-206
3 ; Sequence 206, Application US/09925299
4 Patent No. US20020055627A1
5 ;
6 GENERAL INFORMATION:
7 ;
8 APPLICANT: Rosen et al.
9 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
10 ; FILE REFERENCE: PA102
11 ;
12 CURRENT APPLICATION NUMBER: US/09/925,299
13 ;
14 CURRENT FILING DATE: 2001-08-10
15 ;
16 PRIOR APPLICATION NUMBER: PCT/US00/05883
17 ;
18 PRIOR FILING DATE: 2000-03-08
19 ;
20 PRIOR APPLICATION NUMBER: 60/124,270
21 ;
22 PRIOR FILING DATE: 1999-03-12
23 ;
24 NUMBER OF SEQ ID NOS: 1556
25 ;
26 SOFTWARE: PatentIn Ver. 2.0
27 ;
28 SEQ ID NO 206
29 ;
30 LENGTH: 5145
31 ;
32 TYPE: DNA
33 ;
34 ORGANISM: Homo sapiens
35 ;
36 FEATURE:
37 ;
38 NAME/KEY: misc_feature
39 ;
40 LOCATION: (4)
41 ;
42 OTHER INFORMATION: n equals a,t,g, or c
43 ;
44 NAME/KEY: misc_feature
45 ;
46 LOCATION: (17)
47 ;
48 OTHER INFORMATION: n equals a,t,g, or c
49 ;
50 NAME/KEY: misc_feature
51 ;
52 LOCATION: (5126)
53 ;
54 OTHER INFORMATION: n equals a,t,g, or c
55 ;
56 NAME/KEY: misc_feature
57 ;
58 LOCATION: (5143)
59 ;
60 OTHER INFORMATION: n equals a,t,g, or c
61 ;
62 US-09-925-299-206

```

Query Match	7.8%;	Score 107;	DB 10;	Length 5145;
Best Local Similarity	51.3%;	Pred. No. 2.9e-19;		
Matches 248;	Conservative 0;	Mismatches 235;	Indels 0;	Gaps 0;

Oy		465	ACITCAAGGTCACC	GGGCCCCCGAGGTGCAGAGATCCCAGGACCCC	524
Db		712	ACAGCCCGTGCTC	TGTGTGAAGGTGAACCTGTGCCCTGGTGAATAATGGA	771


```

OY 771 CCTCTGCTTCAGGAGCTTCAGGGACCGTTGGGGAG-----CCTGGGGTGCCTGG 824
    || || || || || || || || || || || || || || || || || || || || ||
Db 1903 CTTGAAGGGGGAACACAGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1962
OY 825 ACCTGGGGGACCTGCGACAGCTTGGGCTTACCAAGGATCCAGGCCCCCAAGGGCCCCC 884
    || || || || || || || || || || || || || || || || || || || || ||
Db 1963 CACGGGTGGGGCCACGGGGCCCCCAGGGGATCCCTGCTCCCTGGAATCAAGGGCCCCC 2022
OY 885 CGGCGCTCTGCGCCATCAGAGGGGATGGTCCCTGCGCCCTCGAATGAGCCCAACCCC 944
    || || || || || || || || || || || || || || || || || || || || ||
Db 2023 GGGGCTCTCCGGGGCCCCCGGGAGACCCCTGTGCTGGGGCCCTTCATGATGAGCTGCAT 2082
OY 945 GGCA 948
    || || || || || || || || || || || || || || || || || || || || ||
Db 2083 CGCA 2086

RESULT 13
US-09-954-531-961
; Sequence 961, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 961
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-961

Query Match 7.7%; Score 106; DB 9; Length 2542;
Best Local Similarity 52.1%; Pred. No. 3.7e-19;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

OY 467 TACAGGTCCACCGGGCCCGAGGGGTTCACAGAGGTGACAGAGATCCACAGGACCCCTG 526
    || || || || || || || || || || || || || || || || || || || || ||
Db 630 TCCTGTGACACCGGGGTCCCGAGGATCAGGGGCTACCCAGGATGAGGAGGAGGAGG 689
OY 527 GCCCACTGCAACAGGACAGAGAGAGAGAGAGAGGGGAGCTTGACACCTGGCCCTG 586
    || || || || || || || || || || || || || || || || || || || || ||
Db 690 GAGAGACGGGCTCTCATGTATTAAGGCAATGGTGGGCGCTATCGGACCTGGGCGCAC 749
OY 587 CGGGTGAAGAGAGCCCAATTTGACAGCTGCTCCCGCCGAGAGCGTGGGGGGAAGGAT 646
    || || || || || || || || || || || || || || || || || || || || ||
Db 750 CGGGTGAAGAGAGTCTTAGGGGACCGCCAGGCTGGGGAGAGAGGAGGAGAGGGA 809
OY 647 CTAAAGGCTCCAGAGCCCAAGGCTCCCGGTTCCTCCCTGGGAAAGCCCGGCTTCAG 706
    || || || || || || || || || || || || || || || || || || || || ||
Db 810 GCCAGATATTCTGTGAGCCCAAGGGATCAGAGGCCCGGAAGAGAGAGAGGAGGAG 869
OY 707 GCCCAGTGGGAGCCAGGCCCCCGGGGCCACACAGGCAAGAGAGACTCCCGGCTTC 766
    || || || || || || || || || || || || || || || || || || || || ||
Db 870 GCATCAACGAGCAAGATGGAGCCCAAGGCAAGGCTGGCATGAGAGGAGTGCAGGACAG 929
OY 767 AGGGCCCTCTGTGCTTCAGGAGCTTCAGGGACCGTTGGGAGAGCTTGGGGTGCCTGGAG 826
    || || || || || || || || || || || || || || || || || || || || ||
```

```

Db 930 CGGACAGCCCCGGAAGTCCAGGCCACCAAGGGCCTAGCGGGTGTGCCAGGCCACGCTGGGA 989
OY 827 CTCGGGAGTCTCCAGAGCTTCTGGGGGTACCAAGCATGACAGGCCGCCCAAGGCCCCCC 886
    || || || || || || || || || || || || || || || || || || || || ||
Db 990 CAAAGAGAGGCCCTCGAGACACAGGATGAGCGGCCCCCGGACAGGGCTTCTGATTTCTG 1049
OY 887 GCGCTCTGCGCCATCAGAGAGCGGTGTGCCCC 919
    || || || || || || || || || || || || || || || || || || || || ||
Db 1050 GTCCCTCTGGGAAAGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1082

RESULT 14
US-09-964-824A-255
; Sequence 255, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 255
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-255

Query Match 7.7%; Score 106; DB 10; Length 2542;
Best Local Similarity 52.1%; Pred. No. 3.7e-19;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

OY 467 TACAGGTCCACCGGGCCCGAGGGGTTCACAGAGGTGACAGAGATCCACAGGACCCCTG 526
    || || || || || || || || || || || || || || || || || || || || ||
Db 630 TCCTGTGACACCGGGGTCCCGAGGATCAGGGGCTACCCAGGATGAGGAGGAGGAGGAGG 689
OY 527 GCCCACTGCAACAGGACAGAGAGAGAGAGAGAGGGGAGCTTGACACCTGGCCCTG 586
    || || || || || || || || || || || || || || || || || || || || ||
Db 690 GAGAGACGGGCTCTCATGTATTAAGGCAATGGTGGGCGCTATCGGAGCCATGGGCGCAC 749
OY 587 CGGGTGAAGAGAGCCCAATTTGACAGCTGCTCCCGCCGAGAGCGTGGGAGAGGAGTGCAGAGGCA 809
    || || || || || || || || || || || || || || || || || || || || ||
Db 750 CGGGTGAAGAGAGTCTTAGGGGACCGCCAGGCTGGGAGAGAGGAGTGCAGAGGAGCA 869
OY 647 CTAAAGGCTCCAGAGCCCAAGGCTCCCGGTTCCTCCCTGGGAAAGCCCGGCTTCAG 706
    || || || || || || || || || || || || || || || || || || || || ||
Db 810 GCCAGATATTCTGTGACCCCAAGGGATCAGAGGCCCGGAAGAGAGAGAGGAGGAGGAG 869
OY 707 GCCCAGTGGGAGCCAGGCCCCCGGGGCCACACAGGCAAGAGAGACTCCCGGCTTC 766
    || || || || || || || || || || || || || || || || || || || || ||
Db 870 GCATCAACGAGCAAGATGGAGCCCAAGGCAAGGCTGGCATGAGAGGAGTGCAGGACAG 929
OY 767 AGGGCCCTCTGTGCTTCAGAGGACTTCAGGACCGTTGGGAGAGCTTGGGGTGCCTGGAG 826
    || || || || || || || || || || || || || || || || || || || || ||
Db 930 CGGGTGAAGAGAGTCCAGGACACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 989
OY 827 CTCGGGAGTCTCCAGAGCTTCTGGGGGTACCAAGCATGACAGGCCCGCAAGGGGCCCCCC 886
    || || || || || || || || || || || || || || || || || || || || ||
Db 990 CAAAGAGAGGCCCTCGAGACACAGGATGAGCGGCCCCCGGACAGGGCTTCTGATTTCTG 1049
OY 887 GCGCTCTGCGCCATCAGAGAGCGGTGTGCCCC 919
    || || || || || || || || || || || || || || || || || || || || ||
Db 1050 GTCCCTCTGGGAAAGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1082
```


Contact: amg@bcm.tmc.edu
Gunnarntre, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRK Plate: 5 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

Location/Qualifiers
1..3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 1031 a 757 c 760 g 757 t
ORIGIN

Query Match 79.1%; Score 1084.6; DB 11; Length 3305;
Best local Similarity 86.9%; Pred. No. 9.5e-257;
Matches 1192; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 1 ATGACACCGCAGCTCACTATTCACAGGTGAGAGACATACCCTATCTCTCA 60
Db 951 ATGACACCGCAGCTCACTATTCACAGGTGAGAGACATACCCTATCTCTCA 60
QY 61 GCCACGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 1011 GCCACGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070
QY 121 GCCATCAAGTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 1071 GCTGTCAAGTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1130
QY 181 ATGATTAGCAATTCAGATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 1131 ATGATTAGCAATTCAGATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1190
QY 241 GAAGTCAGAGCACTGACAGATACCTTACCAACAGAGAGAGAGAGAGAG 300
Db 1191 GATGTTAGAGCACTGACAGATACCTTACCAACAGAGAGAGAGAGAGAG 1250
QY 301 AATAATACCTGGCAACATCCGTTTGATTCGTTCTCTCAGATGCAACAAGATT 360
Db 1251 AATAATACCTGGCAACATCCGTTTGATTCGTTCTCTCAGATGCAACAAGATT 360
QY 361 ATGAGTCGAGGTTAGACACTGAAGTCACTTATCAGTATTTATGGAAGATG 420
Db 1311 ATGAGTCGAGGTTAGACACTGAAGTCACTTATCAGTATTTATGGAAGATG 1370
QY 421 CTGATGACTCCCAAGATGTGCTGATCAATCAATTTTCACTACTACAGAGT 480
Db 1371 CTGATGACTCCCAAGATGTGCTGATCAATCAATTTTCACTACTACAGAGT 1430
QY 481 GGCCCGAGGGGTCAGAGAGTACAGAGATCCAGAGAGAGAGAGAGAGAGAG 540
Db 1431 GGCCCGAGGGGTCAGAGAGTACAGAGATCCAGAGAGAGAGAGAGAGAGAG 1490
QY 541 AAGGAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db 1491 AAGGAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1550
QY 601 CCAATTGAGACAGAGTGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 1551 ACAATTGAGACAGTGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1610

QY 661 GGCCCGAAAGGCTCCGTTGTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 1611 GGTCGCAAGAGATCTGTGTGTGTCCTCCAGAGAGAGAGAGAGAGAGAGAGAG 1670
QY 721 CCAGGCCCCCGGGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 1671 CCAGGAGACACAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1730
QY 781 TTCAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 1731 TTCAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1790
QY 841 GGCTTGCTGGGTGATCCAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 1791 GGCTTGCTGGGTGATCCAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1850
QY 901 TCAGAGAGAGTGTGCTCCGCTGAGAGATGAGAGAGAGAGAGAGAGAGAG 960
Db 1851 TCAGAGAGAGTGTGCTCCGCTGAGAGATGAGAGAGAGAGAGAGAGAGAG 1910
QY 961 GGCTGCCCCGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 1911 GGATGTCCAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1970
QY 1021 GAAATTTTGGAGAGTCAAGAGCTTTCTGTGAAGAGAGAGAGAGAGAGAG 1080
Db 1971 GAAATTTTGGAGAGTCAAGAGCTTTCTGTGAAGAGAGAGAGAGAGAGAGAG 2030
QY 1081 AACACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 2031 AACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2090
QY 1141 ATCGGCTCAGACACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 2091 ATCGGCTCAGACACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2150
QY 1201 GACTACAAAAATTTGAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 2151 GATTACAAAAATTTGAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2210
QY 1261 GAAGACTGTGCTGCTGATTTATGCTGGGAGAGAGAGAGAGAGAGAGAG 1320
Db 2211 GAAGACTGTGCTGCTGATTTATGCTGGGAGAGAGAGAGAGAGAGAGAG 2270
QY 1321 AATACTTCATTTGCGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
Db 2271 AATACTTCATTTGCGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2331

RESULT 2
B0955927 957 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8763247 NIH_MGC_129 Mus musculus cdna IMAGE:6313500
DEFINITION 5', mRNA sequence.

ACCESSION B0955927
VERSION B0955927.1 GI:22371405
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>
AUTHORS 1 (bases 1 to 957)
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: Resgen, Invitrogen Corp.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1M13740 row: 1 column: 13
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES

source

1..957

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6332754"
 /clone_lib="NIH_MGC_130"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Olfactory epithelium; Vector:
 pCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned
 unidirectionally. Primer: Oligo dT. Average insert size
 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this
 is a NIH_MGC library."

BASE COUNT 260 a 269 c 258 g 169 t 1 others
 ORIGIN

Query Match

Best Local Similarity 49.4%; Score 677.8; DB 14; Length 957;
 Matches 780; Conservative 0; Mismatches 113; Indels 5; Gaps 3;

QY 6 CAGCCAGCTCACTCATTCACAGAGTGCAGTGAACATCACCATATCTCTCAGCCAA 65
 DB 13 CCGGAGATCTACGCTCATTCACAGTGCAGTGAACATCACCATATCTCTCAGCCAA 72
 QY 66 CGAGCAGAACCTGAAGAGCTGCAGAGCTTACCAAGATGCAGAGATTAAGAACAGCAT 125
 DB 73 CGAGCAGAACCTGAAGAGCTGCAGAGCTTACCAAGATGCAGAGATTAAGAACAGCAT 132
 QY 126 CAAGTTCACCACTGAGGAGAGCTTCCAGCTCTTGAAGAGCATATTGTGAACATCAT 185
 DB 133 CAAGTTCACCACTGAGGAGAGCTTCCAGCTCTTGAAGAGCATATTGTGAACATCAT 192
 QY 186 TAGCAATATCAGTTACAGAGCCACCACTGCGAGCTGCAGCAGCAATCTAATGATAGT 245
 DB 193 TAGCAATATCAGTTACAGAGCCACCACTGCGAGCTGCAGCAGCAATCTAATGATAGT 252
 QY 246 CAGGAGCACTTGCAGAGATACCTTACCAAGACAGAGATGATCTACCTCTTGAATTA 305
 DB 253 TAGGAGCACTTGCAGAGATACCTTACCAAGACAGAGATGATCTACCTCTTGAATTA 312
 QY 306 TACCTTGGCCCAACATCCGTTTGGATTCTGTTCTCAGAGATCAACAAGATTGTATGAG 365
 DB 313 CACAGTACAGATCCGCTTGGATTCTGTTCTCAGAGATCAACAAGATTGTATGAG 372
 QY 366 GTGAGAGTTAGACACTGAAGTACCACTTATCATGATTTATGAGAAATGAAGTAGT 425
 DB 373 GTCAAGATTAGACACTGAAGTACCACTTATCATGATTTATGAGAAATGAAGTAGT 432
 QY 426 AGAGCTCAGAGATGTCAGCTCATCAAGATTTCATACATCAAGGTCACCGGGCCC 485
 DB 433 TGACTCTCAAGAGGTCAGCTCATCAAGATTTCATACATCAAGGTCACCGGGCCC 492
 QY 486 CAGGGGTCCAAAGAGTGCAGAGAGATCCAGGAGCCCTGGGCCCACTGGCAACAAGAG 545
 DB 493 CAGAGGTCCAAAGAGTGCAGAGAGATCTCAGGAGCACCTGGTCCAACTGGCAACAAGAG 552
 QY 546 ACAGAAAGAGAGAGAGGAGAGCTGTGTCACCTGGCCCTGCGGTGAGAGGAGCAAT 605
 DB 553 ACAGAAAGAGAGAGAGGAGAGCTGTGTCACCTGGCCCTGCGGTGAGAGGAGCAAT 612
 QY 606 TGGACAGAGTGGTCCCGGAGAGAGGTGGCGGCAAGATCTAAGGCTCCCAAGGGCCC 665
 DB 613 TGGACAGAGTGGTCCCGGAGAGAGGTGGCGGCAAGATCTAAGGCTCCCAAGGGCCC 672
 QY 666 CAAAGGCTCCGTTGTTCCCTGGGAAGCCGCGCTTCAAGGCGCCCAAGTGGGAGCCAG 725
 DB 673 CAAAGGATCTCGTGGGTCCCAAGGAGAGCTGGCCCTCAAGAGACTAAGTGGGAGCCAG 732
 QY 726 CCCCGGGGCCCAACAGGAGAGGAGTCCCGGGCCCTCAGGGCTCTGCTTCA 785

DB 733 ACCACAGGTCACACAGAGAGATGAGTCCCTGGCCCTCANGGCCCTCTGCTTCA 791
 QY 786 GGGACTTCAGAGGAGACCGTTGGGAGCGCTGGGAGCTTCGGGAGATG-CCAGGCT 844
 DB 792 GGGACTTCAGAGGAGACCGTTGGGAGCGCTGGGAGCTTCGGGAGATG-CCAGGCT 851
 QY 845 TGGCTGGGAGTACAGCATGCTC--AGGCCCAAGGAGGAGGAGGAGGAGGAGGAGG 899
 DB 852 TGGCAGGGGTCACAGGATGCTCCCTGGGAGGAGGAGGAGGAGGAGGAGGAGG 909

RESULT 3
 BO891432 936 bp mRNA linear EST 16-AUG-2002
 LOCUS BO891432
 DEFINITION AGENCOURT_8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
 5', mRNA sequence.

ACCESSION BO891432
 VERSION BO891432.1 GI:22283446
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Title Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 JOURNAL NIH-MGC <http://img.mgi.nhl.gov/>.
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9abds-r@mail.nih.gov
 Tissue procurement: Mark Macconchie, Ph.D. and Nancy L. Freeman,
 Ph.D.

CDNA Library Preparation: Resgen, Invitrogen Corp.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1M13790 row: n column: 19
 High quality sequence start: 25
 High quality sequence stop: 632.
 Location/Qualifiers

FEATURES

source

1..936

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6332754"
 /clone_lib="NIH_MGC_130"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: olocysts; Vector: pCMV-SPORT6.1.ccd;
 Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
 Primer: Oligo dT. Average insert size 1.95 kb.
 Constructed by Resgen, Invitrogen Corp. Note: this is a
 NIH_MGC library."

BASE COUNT 273 a 250 c 244 g 163 t 6 others
 ORIGIN

Query Match

Best Local Similarity 42.4%; Score 581.2; DB 14; Length 936;
 Matches 667; Conservative 0; Mismatches 106; Indels 4; Gaps 2;

QY 1 ATGAACAGCCAGCTCACTCATTCACAGTGCAGATGAGAGAAATCAACCATATCTCTCA 60
 DB 131 ATGAATATGAGAGTCACTCATTCACAGTGCAGATGAGAGAAATCAACCATATCTCA 190
 QY 61 GCCAAGAGCAGAGACTGAAAGACCTGCAGAGATTTACCAAAATGACAGATATGAACA 120
 DB 191 GCCAAGAGCAGAGACTGAAAGACCTGCAGAGATTTACCAAAATGACAGATATGAACA 250
 QY 121 GCATCAAGTTCACCACTGAGAGAGCTTCCAGCTCTTTGAGAGGATATTGTGAAC 180
 DB 251 GCATCAAGTTCACCACTGAGAGAGCTTCCAGCTCTTTGAGAGGATATTGTGAAC 310
 QY 181 ATCATTTAGCAATATCATTCACAGGCTCAGAGCTGAGCAATCTTAAT 240

Db 311 ATCATTTGCAACATCACTACACAGCCCATCAGTGGAGACATGACACAGATCTTAAT 370
QY 241 GAAGTCAGAGACACCTTCCAGATACCTTACCAAAACACAGATGATGATCTCTCTG 300
Db 371 GATTATGAGACACATGACACAGACCTTGACACAGACAGATGATGATGATGATGATG 430
QY 301 AATTAATCCTTGGCCAACTCCCTTGGATTTCTTTCTCTCAGATGACAAAGATTTG 360
Db 431 AATTAACACTAGTCAACATCCCTTGGATTTCTTTCTCTCAGATGACAAAGATG 490
QY 361 ATGAGTCGAGGTTAGACACTGAGTAGTACCACTATCAGTATGATGAGAAAGATGAG 420
Db 491 ATGAGTCGAGGTTAGACACTGAGTAGTACCACTATCAGTATGATGAGAAAGATGAG 550
QY 421 CTAGTACACTCAGACATGCTGACCTCATCAAGATTTTACATATCTACAGATGACCG 480
Db 551 CTGATGACTCAGACAGAGGTCAGCTCATCAAGATTTTACATATCTACAGATGACCG 610
QY 481 GGGCCCGAGGTCACAGATGACAGAGATCCAGGAGCCCTGACCTCAACTGACAC 540
Db 611 GGGCCCGAGGTCACAGATGACAGAGATCCAGGAGCCCTGACCTCAACTGACAC 670
QY 541 AAGGACAGAAAGAGAGAGAGGAGGAGCTGACACCTGAGGAGGAGAGAGAGG 600
Db 671 AAGGACAGAAAGAGAGAGAGGAGGAGCTGACACCTGAGGAGGAGAGAGAGG 730
QY 601 CCAATGAGACAGCTGCTCCCTCCGAGAGAGCTGAGGAGAGAGATCAAGAGCTCCAG 660
Db 731 ACATTTGACACAGTCCGCTCCGAGAGAGCTGAGGAGAGAGATCAAGAGCTCCAG 790
QY 661 GGGCCCAAGG--CTCCGCTGCTTCCCTTGGAGAGCCGCTGAGGAGGAGGAGG 718
Db 791 GGGTCCAAAGAGATCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 850
QY 719 ACCGAG--CCCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 773
Db 851 ACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 907

RESULT 4
BI456109 906 bp mRNA linear EST 21-AUG-2001
LOCUS 603172765F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5251888 5',
DEFINITION mRNA sequence.
ACCESSION BI456109
VERSION BI456109.1 GI:15246765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
Eumetazoa; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mai.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
[found through the I.M.A.G.E. Consortium/LNL at:](http://image.llnl.gov)
<http://image.llnl.gov>
Plate: LHAM1636 row: b column: 17
High quality sequence stop: 730.
Location/Qualifiers

FEATURES
source 1..906

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5251888"
/clone_lib="NCI_CGAP_Mam5"

/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-Sport6; Site: 1: Saliv;
Site 2: Notti; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lohar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 247 a 228 c 256 g 175 t

Query Match 40.8%; Score 559; DB 13; Length 906;
Best Local Similarity 82.4%; Pred. No. 3,7e-127;
Matches 703; Conservative 0; Mismatches 140; Indels 10; Gaps 5;

QY 224 TGACACAGCAATCTAATAGTAGTACGACCACTTGCAGATATCCCTTACCAACACAG 283
Db 1 TGACACAGCAATCTAATAGTAGTATGATGATGATGATGATGATGATGATGATGATG 60
QY 284 ATGATCTGACCTCTTGAATTAATACCTTGGCCAAACATCCCTTGGATTTCTCTCA 343
Db 61 ATGACCTGACCTCTTGAATTAATACCTTGGCCAAACATCCCTTGGATTTCTCTCA 120
QY 344 GATGACAAAGATTTATAGTTCGAGGTTAGACATGATGATGATGATGATGATGATG 403
Db 121 GATGACAAAGATTTATAGTTCGAGGTTAGACATGATGATGATGATGATGATGATG 180
QY 404 TTATGGAAGAATGAACCTAGTACACTCAAGACATGTCAGCTCATCAAGATTTTACAA 463
Db 181 TTATGGAAGAATGAACCTAGTACACTCAAGACATGTCAGCTCATCAAGATTTTACAA 240
QY 464 TACTACAAAGTTCACCGGGGCCAGAGGTTCCAAAGGTGACAGAGATTCACAGGAGCC 523
Db 241 TTCTACAAAGTTCCTCTGAGGCTCCAGAGGTTCCAAAGGTGACAGAGATTCACAGG 300
QY 524 CTGGCCCACTTGGACACAGAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 583
Db 301 CTGGTCCAACTGGCAACAAAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 584 CTGGGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 643
Db 361 CTGGGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 644 GATCTAAAGGCTCCAGAGGCCCCCAAGAGCTCCGTTGCTCCCTGGAGAGCCGGCCCTC 703
Db 421 GATCTAAAGGCTCCAGAGGCCCCCAAGAGCTCCGTTGCTCCCTGGAGAGCCGGCCCTC 480
QY 704 AGGAGCCCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 763
Db 481 AGGAGCCCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 764 CTGAGGAGGCTCTGCTGCTTCCAGAGGAGCTTACAGGAGAGAGAGAGAGAGAGAGAG 823
Db 541 CTGAGGAGGCTCTGCTGCTTCCAGAGGAGCTTACAGGAGAGAGAGAGAGAGAGAGAG 600
QY 824 GACCTCGGGAGCTGCCAGGCTTGGGCTTACAGGAGAGAGAGAGAGAGAGAGAGAG 883
Db 601 GACCTCGGGAGCTGCCAGGCTTGGGCTTACAGGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 884 CGGAGCTCTGCTGCTTCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943
Db 661 TAGG-CTTCCAGAGGCTTCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 718
QY 944 CGGAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
Db 719 CAGCATAGAGAGGTTCAACAGGAGATGCTGCTACATGGAAGAGATTTACAGGATTAATG 778
QY 999 CTACTATTTTCAAGTTG--AGAAAGAAATTTTGAAGATGACAAAGCTTTCTGTAAGA 1055
Db 779 CTACTATTTTCAATGTCACAAAGAAATTTTGAAGATTTGTAAGGCTTTCTGTAAGA 838
QY 1056 CAAGTCTTCACAT 1068

Db	839	AAATTTTCCAAT	851
RESULT	5		
LOCUS	BO713873		
DEFINITION	BO713873	861 bp	mRNA
ACCESSION	AGNCNCURT_8482828	NIH_MGC_129	Mus musculus CDNA clone IMAGE:6306240
VERSION	5		
KEYWORDS	BO713873		
SOURCE	BO713873.1	GI:21852772	EST.
ORGANISM	house mouse.		
	Mus musculus.		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;		
AUTHORS	Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 861)		
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		

```

Plate: LLM3721 row: n column: 01
High quality sequence stop: 598.
location/Qualifiers
1. .861
FEATURES
source

```

BASE COUNT	ORIGIN
221 a	253 c 239 g 147 t 1 others

	Query Match Best Local Similarity Matches 647/	40.6% 82.8% 0/	Score 556.4/	DB 14/	Length 861/	Mismatches 132/	Indels 2/	Gaps 1/
QY	359	TGATGAGGTCGAGGTTAGACACTGAAAGTAGCCAACTTATCATGATGATTTGAGAGAAATGA						
Db	14	TGATGAGGTCGAGGTTAGACACTGAAAGTAGCCAACTTATCATGATGATTTGAGAGAGATGA						
QY	419	AGCTATGTAGACTCCAGACATGCTCAGCTCATTCAGAAATTTTATCAATTACTACAGSTCCAC						
Db	74	AACGTGTGACATCCAGACACGCTCAGCTCATTCAGAAATTTTATCAATTACTACAGAGTCTCC						
QY	479	CGGGCCCCAGGGGTCGACAGAGGTGCACAGATTCACAGGAGACCCCTTGSGCCCACTGTGCA						
Db	134	CTGGCCCCAGAGGTCGACAAAGGTGCACAGATTCCTCAAGGAGACACCTGTGTCCAACTTGCA						
QY	539	ACAAGGAGCAGAAAGGAGAGAAAGGGGAGCCCTGAGCACCTCGGCCCTCGGGGTGAGAGAG						
Db	194	ACAAGGAGCAGAAAGGAGAGAAAGGAGAGCCGTGATCCACTGTGCCCTCGGGGTGAGAGAG						
QY	599	GCCCAATTGGACACGACTGCTCCCGCCGAGAGACGTGGCGGCAAGAGATCTAAAGGCTCCC						
Db	254	GCACAAATGGAGACGAGTGGCCCTCTCTGGAGAGACGTGGCAGCAAAAGATTCAAAGGCTCAC						
QY	659	AGGGCCCCAAAGGCTCCCGTGGTCCCTCTGGAGAACCCGGGCTCTAAGGGCCCCAATGGGG						
Db	314	AGGGTCCCCAAAGGATCTCTGTGGTCCCTCCAGGAGAACCTGTGGCTCTTAAGGACTTATGGGG						

QY	719	ACCCGAGCCCCCGGGCCACCACGACAAAGAGGACTCCCGGCGCTTACGGGCCCTCTCG	778
Db	374	ACCCAGGACACACAGGTCACCGAGCAAGGATGAGCTCCCTGGCCCTCAGGAGCCCTCTTG	433
QY	779	GCTTCAGGAGCTTCAGGGACACGTTGGGGAGACCTGGGGTCTCTGGACCTGGGGACTGC	838
Db	434	GCTTCAGGAGACTAGGGCACTGGGGTAGACCTGGAGTAGTACCTGGACCTGGGGGCTTGC	493
QY	839	CAGGCTTCCTGGGGTACCAAGGCATGCCAGGCCCAAGGGCCCCCGGGCCCTCTGGCC	898
Db	494	CAGGCTTCACGAGGGGTGCAGAGCAGCTCGGAGCTTAAAGGACACACTGGCCCTTCAGGCC	553
QY	899	CATCAGAGAGCGGTGGTGCCCTGGGCCCTGCAGAAATGAGCCCAACCCCGGACAGGAGACA	958
Db	554	CCTCAGAGAGCAATGAGAGCATTTGGCTCTGCAGAAATGAACCAACCCACATCAGAGTCA	613
QY	959	ATGCGTCCGCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGA	1018
Db	614	ACGGATGTGCGCCTCACTCGAAGAACTTCACAGATAAATGCTACTATTTTTCCTGGCAA	673
QY	1019	AAGAAATTTTGGAGATGTCAAAGC--TTTTCGTGAAGACAGTCTTCACATCTTGTATT	1076
Db	674	AAAAAAATTTTCGAAAAATGCTAACCTTTTTCGCAACAAAAATCTCCCATCCGGGTTT	733
QY	1077	CATAAACTAGAGAGAAACAGACATGATATAAAAAACAGATGTTAGGAGAGAGAGCCA	1136
Db	734	CATAACCCCAAGAAAAACAGCCATGGGATATAAAACCTTACCTTCGGGGAGCAGAAACC	793
QY	1137	C 1137	
Db	794	C 794	

LOCUS	B0674807	668 bp	mRNA	linear	EST 15-JUL-2002
DEFINITION	AGNCOCURT_8035172 NIH_MGC_102 Homo sapiens CDNA IMAGE:62126722				
ACCESSION	5', mRNA sequence.				
VERSION	B0674807				
KEYWORDS	B0674807.1	GI:21785641			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Euthera; Primates; Carnivora; Homnidae; Homo.				
JOURNAL	1 (bases 1 to 668)				
COMMENT	NIH-MGC http://mgc.ncl.nih.gov/ .				
	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgabbs@remail.nih.gov				
	Tissue Procurement: ATCC				
	CDNA Library Preparation: Rubin Laboratory				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNL at:				
	http://image.lnl.gov				
	Plate: LICM2377 row: c column: 09				
FEATURES	High quality sequence stop: 656.				
source	Location/Qualifiers				
	1..668				

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:6212672"
/clone_1fp="NH.MGC.102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: salivary gland; Vector: pCM7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G) Library constructed by Ling Hong in the laboratory of Gerald M. Rubin

```

```
/clone_lib="LTI_FL002_PL1"
```

Email: cyapbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM9128 row: m column: 13
High quality sequence stop: 580.

FEATURES

Source

Location/Qualifiers

1..580

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3962292"

/clone_lib="NCI-GCAP_Mam1"

/tissue_type="tumor, biopsy sample"

/dev_stage="10 months, virgin"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt;

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT

159 a 153 c 161 g 107 t

ORIGIN

Query Match

32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 85.5%; Pred. No. 3.1e-99;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 635 GCGGCAAGATCTAAAGCTCCAGGGCCCAAGAGCTCCGTTGTTCCCTGGAGAC 694

Db 1 GCAGCAAGATGCAAGAGCTCAGAGGTCCTCAAGATCTGTGTGTCCTCCAGGAGAC 60

QY 635 CCGGCTTCAGAGGCGCCAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAGAGAGAC 754

Db 61 CTGGCGCTCAAGAGACCTGTGGGAGCCAGGACCCAGGACAGGTCACAGGCAAGAGATGAC 120

QY 755 TCCCGGCTCAGAGGCTCTGCTGCTCCAGGACTTCAGGACCTGTTGGGAGACCTG 814

Db 121 TCCGCGCTCAGAGGCTCTGCTGCTCCAGGACTTCAGGACCTGTTGGGAGACCTG 180

QY 815 GGGTCTGAGACTGCGGAGCTCCAGGCTTCCTGCTGCTACAGGACATGCCAGGCCCA 874

Db 181 GAGTACCTGAGACTGCGGAGCTTCAGGCTTCAGGAGGATGCCAGGACATGCCAGGCCCA 240

QY 875 AGGGCCCCCGGCTCTGCTGCTCCAGGACGAGGCTGCTGCTGCTGCTGCTGCTGCTG 934

Db 241 AGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 935 AGCCAAAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 994

Db 301 AACCAAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360

QY 995 AATGCTACTATTTTTCAGTGGAGAAATTTTGGAGGCAAAATTTTCTGCTGCTGCTG 1054

Db 361 AATGCTACTATTTTTCAGTGGAGAAATTTTGGAGGCAAAATTTTCTGCTGCTGCTG 420

QY 1055 ACAAGCTTTCACATCTGTTTTCATTAACATAGAGAGGAGGAGGAGGAGGAGGAGG 1114

Db 421 ACAAGCTTTCACATCTGTTTTCATTAACATAGAGAGGAGGAGGAGGAGGAGGAGG 480

QY 1115 AGATGCTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1174

Db 481 AATGCTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540

QY 1175 GGAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1214

Db 541 GGAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 580

RESULT 9

AW958053

LOCUS

AW958053

552 bp

mRNA

linear

EST 01-JUN-2000

DEFINITION

EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.

ACCESSION

AW958053

VERSION

AW958053.1 GI:8147736

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS

Hedge, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.

TITLE

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

JOURNAL

Unpublished (2000)

COMMENT

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: john@igf.org
Plate: 115

FEATURES

Seq primer: Reverse.

Source

Location/Qualifiers

1..552

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGE"

/note="Vector: pBluescriptSk"

BASE COUNT

176 a 112 c 143 g 121 t

ORIGIN

Query Match

31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 99.1%; Pred. No. 3.6e-97;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 919 CTGGCCCTCAGAGTGGAGCAACCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 978

Db 1 CTGGCCCTCAGAGTGGAGCAACCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60

QY 979 AAGAACTTCAGAGCAAAATGCTACTATTTTTCAGTTGAGAGAAATTTTGGAGATGA 1038

Db 61 AAGAACTTCAGAGCAAAATGCTACTATTTTTCAGTTGAGAGAAATTTTGGAGATGA 120

QY 1039 AAGCTTTCTGAGAGCAAAATGCTACTATTTTTCAGTTGAGAGAAATTTTGGAGATGA 1098

Db 121 AAGCTTTCTGAGAGCAAAATGCTACTATTTTTCAGTTGAGAGAAATTTTGGAGATGA 180

QY 1099 CAATGATTAATAAAGAGAGTGGAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 1158

Db 181 CAATGATTAATAAAGAGAGTGGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 240

QY 1159 GAGCGTGAATAAGATGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1218

Db 241 GAGCGTGAATAAGATGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 300

QY 1219 GCTGACAGCCGAGTAACTGGGGTCAATGGCCATGAGGAGGAGGAGGAGGAGGAG 1278

Db 301 GCTGACAGCCGAGTAACTGGGGTCAATGGCCATGAGGAGGAGGAGGAGGAGGAG 360

QY 1279 ATTATGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1337

Db 361 ATTATGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420

QY 1338 AAAAGACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1371

Db 421 AAAAGACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 454

RESULT 10

BE290299

LOCUS

BE290299

601 bp

mRNA

linear

EST 26-OCT-2000

DEFINITION	601089246p1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5'
ACCESSION	mRNA sequence. BE290299
VERSION	BE290299.1 GI:9171250
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 601)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLM8516 row: k column: 12 High quality sequence stop: 551. Location/Qualifiers
FEATURES	source 1..601
	/organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:3484163" /clone_1lb="NCI_CGAP_Mam5" /tissue_type="tumor, gross tissue" /dev_stage="7 months" /lab_host="DH10B" /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
BASE COUNT	156 a 161 c 179 g 105 t
ORIGIN	
Query Match	30.3%; Score 415.8; DB 10; Length 601;
Best Local Similarity	84.8%; Pred No. 7.3e-92;
Matches 501; Conservative	0; Mismatches 87; Indels 3; Gaps 3
Db	541 AAGGCACAAAGAGAAGAGGGGAGCCCTGGACCACTCGGCCGTGGTGAGAGAGGC 600 9 AGGTGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 68
Db	601 CCAATTGGACCAAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAGGCTCCAG 660 69 ACAATTGGACCAAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAGGCTCCAG 128
Db	661 GGCCCCAAAGGCTCCCGGTGGTTCCCTGGG -AAGCCCGGCCCTCAGAGGCGCCAGTGGGGA 719 129 GGTTCCCAAAGGATCTGTGTGGTGTCCCAAGGATTAAGCTGTGCTTCAGAGACTAAGTGGGA 188
Db	720 CCCAGGCCCCCGGGGCCACCAGAGCAAAAGAGGACTCCCGGCCCTCAGAGGCTTCCTGG 779 189 CCCAGAGCACACAGGCTCCACAGGACAGAGATGAGATCTCTGGCCCTCAGAGGCTTCCTGG 248
Db	780 CTTCACAGGACTTCAGAGGACCGCTTGGGAGGCTGTGGGCTCTGAGACTCTGGGGACTGCG 839 249 CTTCACAGGACTTCAGAGGACCGCTTGGGAGGCTGTGGGCTCTGAGACTCTGGGGAGTGGC 308
Db	840 AGGCTTGGCTGGGGATACACAGGATGACAGGCCCCCAAGGGCCCCCGGCCCTTCGTGGCCC 899 309 AGGCTTGGCC -AGGTTGGCCAGGACATGCTTGGGCTTAAGGAGCAACCTGCGCCCTTCAGAGCCC 367
Db	900 ATCAGAGAGGAGGTGGTCCCTCTGGCCCTGTGAGAAATGAGCAACCCCGGACACGGAGAGACA 959 368 CTCAGAGAGAAATGGAGCCATTGGCTCTTGAGATGAACACCCAGCATCAGAGAGGTCAA 427

QY	960	TGCGTGGCCGCTCAGCTCGGAGAACTTCACAGACAAATGCTACTTTTTCAGTTGAGAA	1019
Db	428	CGGATGTCGGCTCCTCTGGAAGAACTTCACGATTAATGCTACTTTTTCAGTTGAGAA	487
QY	1020	AGAAATTTTTCAGAGTTCGAAAGGCTTTTTCGTGAGACAGCTCTTCACATCTTGTTCAT	1079
Db	488	AGAAA-TTTTGAAGATGCTGAAGCTTTTCTGTGAGACAAATCTTCCATCTCTGTTTCA	546
QY	1080	AAACACTGAGAGACAGCAATGATATAAAAAACAGATGCTAGGAGAGA	1130
Db	547	AAACTCAAGAGAGAAACAGCAATGCTATAAAAAACGATACGTTGGGAGAGA	597
RESULT 11			
LOCUS	B0934501	884 bp	mRNA linear EST 21-AUG-2002
DEFINITION	AGENCOURT.8/65665 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330794		
ACCESSION	B0934501		
VERSION	B0934501.1	GI:22349884	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 884)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@btl-rt@mail.nih.gov Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman, Ph.D.		
FEATURES			
source	CDNA Library Preparation: ResGen, Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov plate: LLAM13785 row: m column: 03 High quality sequence stop: 352. Location/Qualifiers 1..884		
	/organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:6330794" /clone_lib="NIH_MGC_130" /lab_host="DH10B (phage-resistant)" /note="Organ: otcysts; Vector: pcMV-SpORF6.1.ccd; Site:1: EcoRV; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library"		
BASE COUNT	282 a	225 c	205 g 172 t
ORIGIN	28.1%; Score 385.6; DB 14; Length 884; Best Local Similarity 82.9%; Pred. No. 2.5e-84; Matches 489; Conservative 0; Mismatches 94; Indels 7; Gaps 4;		
QY	1	ATGACAGCCAGCTCACTCATTTACAA-GGTGAGATGAGAAATCACCAGTATCTTCA	59
Db	277	ATGATATACCAAGCTCAGCTCATTTACACAGGCTGAGATGACAACTATACCATATCTTCA	336
QY	60	AGCCACAGACGAGCAACCTGGAAGACCTGACAGACTTACACAAAGATGCAGAAATATGAA	119
Db	337	GGCCAAGACGAGCAAGCTTGAAGACCTTCAGGACTTACACAAAGATATGAGAAATATGAA	396
QY	120	AGCCATCAAGTTTCAACCAACTGAGAGAAAGCTTCAGGCTTTGAGACGATATTTGTGA	179
Db	397	AGCTGTCAAGTTTCAAGCCCAACTTGAAGAAAGCTTCAAGGCTTTTGAGACAGATATTTGTGA	456

Y	180	CATCATTTGACATATTCAGTTTACACAGCCACACCGCTGGGACGCGACCAATCTMAA	239
Db	457	CATCATTTGACATATTCAGTTTACACAGCCACACCGCTGGGACGCGACCAATCTMAA	516
Y	240	TGCAATCTAGGACACCATCTTGCACAGATACCTCTTACCAAAACACAGATGATCTGACCTCTT	299
Db	517	TGATTTTGGAGACCATGACAGACACCTTGACACAGACAGATGATGATCTGACCTCTT	576
Y	300	GAAATACCATCTTGGCCACATCCGTTTGGATTTCTGTTTCTGACAGATGCAACAGATTT	359
Db	577	GAAATACCATCTTGGCCACATCCGTTTGGATTTCTGTTTCTGACAGATGCAACAGATTT	635
Y	360	GATGAGGTCGAGGTTTACACATCTGAGTACGACCACTTATCAGTATGGAAGAATGAA	419
Db	636	GATGAGGTCGAGGTTTACACATCTGAGTACGACCACTTATCAGTATGGAAGAATGAA	695
Y	420	GCTATATACATCTTCCAGCATCTGCTGACCTGATCAGAAATTTTACAACTACAGAGTCCAC	479
Db	696	ACTGTTTGCATCTTCCAGCATCTGCTGACCTGATCAGAAATTTTACAACTACAGAGTCCAC	755
Y	480	GGGCCCCGAGGGTCCAGAGGTG--ACAGAGGATGCCAGGAGCCCTGTG--CCCAACT	534
Db	756	TGGCCCCGAGGGTCCAGAGGTG--ACAGAGGATGCCAGGAGCCCTGTG--CCCAACT	815
Y	535	GGCAACAAGGAGACAGAAAGGAGAGAGGGGGAGCCTGTGACCACTGCGCC	584
Db	816	GGCAACAAGGAGAGACAAAGGAGAGAGGGGGAGCCTGTGCTC	865
RESULT 12			
LOCUS	BB248064	638 bp	mRNA
DEFINITION	BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus		
ACCESSION	BB248064		
VERSION	BB248064.2	GI:16355610	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murine; Mus.		
AUTHORS	1 (bases 1 to 638)		
	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,T., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakawa,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.		
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)		
JOURNAL	Unpublished (2001)		
COMMENT	On Jul 6, 2000 this sequence version replaced gi:8940810. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shenocho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp/ URL://genome.gsc.riken.go.jp/ waghi,K., Fujiwara,S., Inoue,K., Togawa,Y., Iwata,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Okawa,K., Tanaka,T., Matsura,S., Kawai,T., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multiplexed capillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)		

Query Match	Best Local Similarity	Score	DB	Length	658;
Matches	415;	Conservative	0;	Mismatches	76;
				Indels	0;
				Gaps	0;
OY	881	CCCCGGGCGCCTGCGCCATCAGGAGGAGGGGTCGCCCTGGCCCTGCAAGATGAGCCAA	940		
Db	1	CACGTGGCCCTCCAGGCCCTCAGGAGCAATGAGCCCAATTGGCTTTCGAGATGTAACCA	60		
OY	941	CCCCGGCACCGGAGGACATGGCTCCCGCCCTCAGCTGGAAGAATCTTCACAGCAAAATGCT	1000		
Db	61	CCCCGACATCAGAGGTCGAACGATGTCGCGCTCATGTGAAGAATCTTCACAGTAATATGCT	120		
OY	1001	ACTATTTTTCAGTTGAGAAAGAAATTTTTCAGAGATGCAAAAGCTTTTCTGTGAGACAAAGT	1060		
Db	121	ACTATTTTTCATTTGGAAGAAAGAAATTTTTCAGAGATGCAAAAGCTTTTCTGTGAGACAAAT	180		
OY	1061	CTTCACATCTGTTTTCATTAACACTAGAGAGGAAGACCAATGGATTAATAAAAGCAATGG	1120		
Db	181	CTTCACATCTGTTTTCATTAACACTAGAGAGGAAGACCAATGGATTAATAAAAGCAATACG	240		
OY	1121	TAGGAGAGAGAGACCATGATCGGCTTCACAGATCAGACGTCGTAATAATGATGGAAGT	1180		
Db	241	TGGGAGAGAAAGCATTTGATCGGCTTCACAGATCAGACGTCGTAATAATGATGGAAGT	300		
OY	1181	GCGTGGATGGACATCTCCAGACTACAAAAATTTGGAAGCTTGAGACGCGGATTAATGGG	1240		

Db 301 GGGTACAGGGGTCACCTGTTGATTACAAAGCTGAGAACCCAGATCACTGGG 360
 QY 1241 GTCATGCGCCATGGCCAGAGAGACCTGCTGCGGTTGATTTATGCTGGCAGTGAAGC 1300
 Db 361 GGAGTGGCCATGGCCAGAGAGACCTGCTGCGGTTGATTTATGCGAGAGACAGTGAAGC 420
 QY 1301 ATTTCGAATGCAAGACGTCAATTAATCTTATTTGCGAAAAAGACGAGACAGTACTGT 1360
 Db 421 ACTTCAGATGATGAATCAATTAATCTTATTTGCGAAAAAGACGAGAGAGTACCAT 480
 QY 1361 CATCTGACATTA 1371
 Db 481 CATCCATATTA 491

RESULT 13
 BM676508/c 500 bp mRNA linear EST 27-FEB-2002
 LOCUS BM676508 UI-E-EJ0-abq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone
 DEFINITION UI-E-EJ0-abq-c-05-0-UI 3', mRNA sequence.
 ACCESSION BM676508
 VERSION BM676508.1 GI:18986404
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 500)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward

FEATURES

source

Location/Qualifiers

1..500
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-abq-c-05-0-UI"
 /clone_lib="UI-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT733-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into the pT733-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAGA
 ; lens, CGATTACGCA; eye anterior segment, AATGCCGCAT;
 optic nerve, CCATTACG; retina, CCGC; Retina foveal and

Macular, GTC; RPE and Choroid, ACCCA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI).
 TAG_LIB=UI-E-EJ0
 TAG_TISSUE=human eye anterior segment
 TAG_SEQ=AAATGCCCAT"
 BASE COUNT 109 a 123 c 89 g 179 t
 ORIGIN

Query Match 25.6%; Score 351.4; DB 14; Length 500;
 Best Local Similarity 99.7%; Pred. No. 5.4e-76;
 Matches 352; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1019 AAGAAATTTTGGAGATGCAAGACCTTTCTGTGAACAAGCTTTCACATCTGTTTCA 1078
 Db 494 AGGAATTTTGGAGATGCAAGACCTTTCTGTGAACAAGCTTTCACATCTGTTTCA 435
 QY 1079 TAAACACTAGAGAGAGACGCAATGATTAATAAACAGTGTAGGAGAGAGCCACT 1138
 Db 434 TAAACACTAGAGAGAGACGCAATGATTAATAAACAGTGTAGGAGAGAGCCACT 375
 QY 1139 GGATCGGCTCCACAGACTGAGACCGTGAATAATGAATGAGTGGCTGATGGACATCTC 1198
 Db 374 GGATCGGCTCCACAGACTGAGACCGTGAATAATGAATGAGTGGCTGATGGACATCTC 315
 QY 1199 CAGACTACAAAATTTGGAAGCTGACACCCCGATTAATGCGGCTCATGGGCCAG 1258
 Db 314 CAGACTACAAAATTTGGAAGCTGACACCCCGATTAATGCGGCTCATGGGCCAG 255
 QY 1259 GGAACAGCTGCTGGGTGATTTATGCTGGCAGTGAACGATTTCCAAATGTGAAGAG 1318
 Db 254 GGAACAGCTGCTGGGTGATTTATGCTGGCAGTGAACGATTTCCAAATGTGAAGAG 195
 QY 1319 TCAATTAATCTTATTTGCGAAAAAGACAGAGAGACACTACTGTCATCTGATTA 1371
 Db 194 TCAATTAATCTTATTTGCGAAAAAGACAGAGAGACACTACTGTCATCTGATTA 142

RESULT 14
 LOCUS BM713891

DEFINITION UI-E-EJ0-abq-c-05-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-abq-c-05-0-UI 5', mRNA sequence.

ACCESSION BM713891
 VERSION BM713891.1 GI:19027149
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 507)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 443-498, >POLY_ASimple_repeat
 Seq primer: M13 Reverse.
 Location/Qualifiers

FEATURES

```

source
1. 507
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="UI-E-EJ0-abq-c-05-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-E-EJ0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAGCA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAAT;
optic nerve, CCATTAACTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
```

BASE COUNT 184 a 89 c 121 g 112 t 1 others

ORIGIN

Query Match 25.5%; Score 350; DB 14; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.2e-75;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1022 AAATTTTGAGAGTCAAAAGCTTTCTGTGAAGACAGCTTTCATCTTTTCATTA 1081
|||||
Db 1 AAATTTTGAGAGTCAAAAGCTTTCTGTGAAGACAGCTTTCATCTTTTCATTA 60

OY 1082 ACACATAGAGAGAGACAGCAATGATATAAAACAGATGTAGAGAGAGACCCACTGGA 1141
|||||
Db 61 ACACATAGAGAGAGACAGCAATGATATAAAACAGATGTAGAGAGAGACCCACTGGA 120

OY 1142 TCGGCTCAGACAGCTCAGAGCGTGAATGATGAGAGTGGCTGAGACATCTCCAG 1201
|||||
Db 121 TCGGCTCAGACAGCTCAGAGCGTGAATGATGAGAGTGGCTGAGACATCTCCAG 180

OY 1202 ACTACAAAAATTGGAAGCTGACACCGGATTAAGTGGGTCATGGCCATGGCCAGAG 1261
|||||
Db 181 ACTACAAAAATTGGAAGCTGACACCGGATTAAGTGGGTCATGGCCATGGCCAGAG 240

OY 1262 AAGACTGTGCTGGTGAATTTATGCTGGGCACTGGAAAGATTCATGTAAGAGCTCA 1321
|||||
Db 241 AAGACTGTGCTGGTGAATTTATGCTGGGCACTGGAAAGATTCATGTAAGAGCTCA 300

OY 1322 ATAACTTCAATTTGCGAAAAAGACAGGAGACAGTACTGTCACTGCAATTA 1371
|||||
Db 301 ATAACTTCAATTTGCGAAAAAGACAGGAGAGACAGTACTGTCACTGCAATTA 350

RESULT 15
B0771366 808 bp mRNA linear EST 26-JUL-2002
LOCUS B0771366
DEFINITION UI-M-F10-byu-9-09-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
IMAGE:5702432 5', mRNA sequence.
ACCESSION B0771366
VERSION B0771366.1 GI:21979842
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.

FEATURES
source
Location/Qualifiers
1. 808
/organism="Mus musculus"
/strain="C57BL/6"
/db.xref="taxon:10090"
/clone="IMAGE:5702432"
/tissue_type="Whole Brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: Ecor I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CAGCCACGAC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
Developing Mouse Nervous System", supported by National
Institute of Mental Health (NIMH), Hemlin Chhn, Ph.D.,
program coordinator."

BASE COUNT 270 a 158 c 173 g 206 t 1 others

ORIGIN

Query Match 25.1%; Score 343.8; DB 14; Length 808;
Best Local Similarity 85.1%; Pred. No. 5e-74;
Matches 384; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 921 GGCCTCGAGAAATGAGCCACCCCGACCGAGAGCAATGCGTCCGCTCACTGGA 980
|||||
Db 1 GGCCTCGAGAAATGAGCCACCCCGACCGAGAGCAATGCGTCCGCTCACTGGA 60

OY 981 GAACCTTCACAGCAAAATGTACTATTCTTCAGTGAAGAAATTTTGAGATGCCAA 1040
|||||
Db 61 GAACCTTCACAGCAAAATGTACTATTCTTCAGTGAAGAAATTTTGAGATGCCAA 120

OY 1041 GCTTTTCTGGAAGCAATCTTCAATCTTTTTCATTAACACTAGAGAGACAGA 1100
|||||
Db 121 GCTTTTCTGGAAGCAATCTTCCATCTCGTTTTCATTAACACTAGAGAGACAGA 180

OY 1101 ATGAGTAAAAAAGACAGATGTAAGAGAGAGACCACTGGATCGGCTCACTGACA 1160
|||||
Db 181 ATGAGTAAAAAAGACATACCGTGGGAGAGAAAGCAATTTGATCGGCTCACTGACA 240

OY 1161 GGGTGAATTAATGATGAATGCTGATGAGATGAGCAATCTCAAGCTACAAAAATTTGGAAC 1220
|||||
Db 241 ACAGGAACCGAATGGAATGCTGATGAGCGGTCACCTGTGATTACAAAAACTGGAAAC 300

OY 1221 TGGACAGCCGGAATTAATGCGGTATGAGCCATGGGCGAGAGAAAGTGTGCTGGTTGAT 1280
|||||
Db 301 TGGACAGCCAGATTAATGCGGTATGAGCCATGGGCGAGAGAAAGTGTGCTGGTTGAT 360

OY 1281 TTATGCTGGCGATGAGACGATTTCCATATGTAAGACGTCATTAATCTTATTTGCAAAA 1340
|||||
Db 361 TTACGAGAGACAGATGGAATGAGCTTCAGTGTGATGAATTAATCTTATTTGTGAGAA 420

QY 1341 AGACAGGAGACAGTCTCATCTGCATTA 1371
|| ||||| ||||| ||||| |||||
DB 421 GGAAGGAGGAGCAGTACCATCATCATATTA 451

Search completed: March 21, 2003, 08:08:10
Job time : 1811.64 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:02:25 ; Search time 2598.25 Seconds

(without alignments)
11492.169 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_1695

Perfect score: 1026

Sequence: 1 atgcaacaagattgacgag.....cagtaactgcatcgcattta 1026

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_com.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rnd.*
36: em_hcg_mam.*
37: em_hcg_vit.*
38: em_sy.*
39: em_hcg_hum.*
40: em_hcg_mus.*
41: em_hcg_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1026	100.0	2983	9	AB005145
2	1024.4	99.8	1886	9	HS0802985
3	1024.4	99.8	2005	6	AX454442
4	1024.4	99.8	2005	6	AX490920
5	1024.4	99.8	2641	6	AX047353
6	1024.4	99.8	3058	9	AB038518
7	802	78.2	2637	10	AB078434
8	800.4	78.0	3291	10	AB038519
9	615.4	60.0	4330	9	AB052103
10	488.8	47.6	169088	2	AC016128
11	488.8	47.6	188439	9	AP000915
12	484.6	47.2	71044	2	AC024368
13	466.2	45.4	187635	2	AP001022
14	465.8	45.4	178022	2	AP000900
15	437.4	42.6	188255	2	AP000939
16	388	37.8	182029	2	AC114677
17	388	37.8	193208	2	AC102618
18	378.4	36.9	130763	2	AC112416
19	157.6	15.4	3636	9	AB007829
20	157.6	15.4	3685	6	E32511
21	157.6	15.4	3810	6	E32509
22	150.6	14.7	2215	10	BC026446
23	146.8	14.3	178022	2	AP000900
24	146.8	14.3	187635	2	AP001022
25	129	12.6	6156	10	AB009993
26	129	12.6	188255	2	AP000939
27	128.2	12.5	2193	3	PALCOLIA
28	125.8	12.3	6076	10	AF272661
29	124.2	12.1	1041	4	AF138883
30	124.2	12.1	2053	5	AB008374
31	124.2	12.1	4502	5	AB052836
32	122.6	11.9	4995	10	MWU16789
33	121.6	11.9	810	6	AR014116
34	121.4	11.8	4628	4	AB008683
35	121.4	11.8	5676	6	E07265
36	121.4	11.8	5676	9	HUMCALV
37	121.4	11.8	7138	9	HUMPAIV
38	121.4	11.8	4581	4	EC62528
39	119.8	11.7	2010	9	HUMAX2XICOL
40	119.4	11.6	5551	10	AF272662
41	118.4	11.5	6109	6	AX239611
42	118.4	11.5	6109	10	AF176645
43	118	11.5	756	6	AR014110
44	118	11.5	6114	10	CRUPAIV
45	118	11.5	130763	2	AC112416

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AB005145	AB005145	2983 bp mRNA	linear	PRI 21-NOV-2001					
AB005145	AB005145	Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.							
AB005145.1	GI:17026100								
		Homo sapiens female tissue_11b:placenta cDNA to mRNA.							
		Homo sapiens							
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
		Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,							
		Sakai,Y., Fukuoh,A., Sakamoto,T., Itabe,H., Suzutani,T.,							
		Ogasawara,M., Yoshida,I. and Wakamiya,N.							

Pred. No. is the number of results predicted by chance to have a


```
source
1. .1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFZp547G1215"
/lisue_type="brain"
/clone_id="547 (synonym: hfbri). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="fetal"
1. .1886
/gene="DKFZp547G1215"
<1. .1256
/gene="DKFZp547G1215"
/notes="similarity to asialoglycoprotein receptor (Mus
musculus"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKENOLEERFQLEETDVIINISYTAHHLRTLSNLEVRT
TCTDTLTKHTDLSLNNTLANIRDSVLSMODIMSRIDTEVANLSVMEKIV
DSKHGILIKNTIIGPPGPRGPRGDSOGPRPTGKDGKGEKGEPRGPRGERS
PIGAPRPRGEGKSGSKSGSKSGSKSGSKSGSKSGSKSGSKSGSKSGSKSG
PFGGLGTVGEPVPGRLPLPGVPMGPFGPPGPPGSAVPLALQNEPTPA
PEDNGCPPHMKNFTDKCYFSEKEIFEDATFCEDKSHLVFINTREOQIKOMY
GRESHWIGLIDSERENEMKWLDTSPDYKNMKAGQPDNMGHGHPGDEAGLIYAGCW
NDFCEVDNNFICEKRETVLSAL"
1838. .1843
polyA_signal
/gene="DKFZp547G1215"
1869
polyA_site
/gene="DKFZp547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Query Match 99.8%; Score 1024.4; DB 9; Length 1886;
Best Local Similarity 99.9%; Pred. No. 8.7e-226;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 541 GGCCTCTCGGCCATCAGAGAGCGGCGGCCCTGGCCCTGCAGATGAGCCAAACCGC 600
|||||
DB 768 GGCCTCTCGGCCATCAGAGAGCGGCGGCCCTGGCCCTGCAGATGAGCCAAACCGC 827
|||||
QY 601 GCACCGAGAGACAAATGCTGCGCCCTCCTCAGTGAAGAACTTGCAGACAGCAATGCTACTAT 660
|||||
DB 828 GCACCGAGAGACAAATGCTGCGCCCTCCTCAGTGAAGAACTTGCAGACAGCAATGCTACTAT 887
|||||
QY 661 TTTTCAGTTGACAAAGAAATTTTGTGAGATGCAAAAGCTTTTGTGAGACAAAGTCTTGA 720
|||||
DB 888 TTTTCAGTTGACAAAGAAATTTTGTGAGATGCAAAAGCTTTTGTGAGACAAAGTCTTGA 947
|||||
QY 721 CATCTGTTTATTAACCTAGAGAGCAACGATGATTAATAAACAAGATGATGAGG 780
|||||
DB 948 CATCTGTTTATTAACCTAGAGAGCAACGATGATTAATAAACAAGATGATGAGG 1007
|||||
QY 781 AGAGAGAGCCACTGATGCGCCTCAGACAGCTGAGACCGTGAATAATGAATGAAGTGGCTG 840
|||||
DB 1008 AGAGAGAGCCACTGATGCGCCTCAGACAGCTGAGACCGTGAATAATGAATGAAGTGGCTG 1067
|||||
QY 841 GATGGACATCTTCAGACTACAAAAATTTGAAAGCTGGACACCCGATTAATGGGGTCA 900
|||||
DB 1068 GATGGACATCTTCAGACTACAAAAATTTGAAAGCTGGACACCCGATTAATGGGGTCA 1127
|||||
QY 901 GGCATGGGCGCAGAGAAAGCTGCTGGGTGATTTATGCTGGGACATGGAACATTTTC 960
|||||
DB 1128 GGCATGGGCGCAGAGAAAGCTGCTGGGTGATTTATGCTGGGACATGGAACATTTTC 1187
|||||
QY 961 CAATGTGAAGACGTCAATTAATTCATTTGCGAAAAAGACAGAGACAGTACTGTCACT 1020
|||||
DB 1188 CAATGTGAAGACGTCAATTAATTCATTTGCGAAAAAGACAGAGACAGTACTGTCACT 1247
|||||
QY 1021 GCATTA 1026
|||||
DB 1248 GCATTA 1253
|||||

RESULT 3
AX454442
LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 27 from Patent WO208284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A.,
Gadowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I.
and Ye,W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,
Audrey (US) ; Gadowski, Paul J. (US) ; Gurney, Austin L. (US) ;
Hillan, Kenneth J. (US) ; Marsters, Scott A. (US) ; Pan, James (US)
; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
I. (US)
FEATURES
source Location/Qualifiers
1. .2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Query Match 99.8%; Score 1024.4; DB 6; Length 2005;
Best Local Similarity 99.9%; Pred. No. 8.7e-226;
```

Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1	ATGCACACAGATTGATGAGTGCAGGTTGAGACACTGAACTAGCAACATTCAGTGTAT	60	
DB	478	ATGCACACAGATTGATGAGTGCAGGTTGAGACACTGAACTAGCAACATTCAGTGTAT	537	
QY	61	ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACATA	120	
DB	538	ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACATA	597	
QY	121	CTTCAAGTTCACACGGGGCCCCAGGGTCCCAAGAGTGCAGAGAGTTCACAGGACCCCT	180	
DB	598	CTTCAAGTTCACACGGGGCCCCAGGGTCCCAAGAGTGCAGAGAGTTCACAGGACCCCT	657	
QY	181	GGCCCAACTGGCAACAAGGACAGAAAGAGAGAAGGGAGGAGCTGGACCACTGGCCCT	240	
DB	658	GGCCCAACTGGCAACAAGGACAGAAAGAGAGAAGGGAGGAGCTGGACCACTGGCCCT	717	
QY	241	GGGGGTGAGAGAGGCCCAATTGGACAGCTGTCCCGCCGAGAGCGTGGCGCAAGA	300	
DB	718	GGGGGTGAGAGAGGCCCAATTGGACAGCTGTCCCGCCGAGAGCGTGGCGCAAGA	777	
QY	301	TTCTAAGGCTCCAGGGCCCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGCCCTCAG	360	
DB	778	TTCTAAGGCTCCAGGGCCCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGCCCTCAG	837	
QY	361	GGCCCAAGTGGGAGACCCAGGCCCCCGGGCCCAACAGGACAGGAGGAGCTCCCGGCCCT	420	
DB	838	GGCCCAAGTGGGAGACCCAGGCCCCCGGGCCCAACAGGAGGAGCTCCCGGCCCT	897	
QY	421	CAGGGCCCTCTGGCTTCCAGGAGCTTTCAGGACCGCTGGGAGCCCTGGGCTCTGGA	480	
DB	898	CAGGGCCCTCTGGCTTCCAGGAGCTTTCAGGAGCGCTGGGAGCCCTGGGCTCTGGA	957	
QY	481	CTTCGGGGAGTGCAGGCTGCTGGGGTTCACAGGATGCGAGGCGCCCAAGGGCCCCC	540	
DB	958	CTTCGGGGAGTGCAGGCTGCTGGGGTTCACAGGATGCGAGGCGCCCAAGGGCCCCC	1017	
QY	541	GGCCCTCTGGGCCATCAGAGAGCGGTGTCGCCCTGGCCCTGCAGATGAGCCAAACCCG	600	
DB	1018	GGCCCTCTGGGCCATCAGAGAGCGGTGTCGCCCTGGCCCTGCAGATGAGCCAAACCCG	1077	
QY	601	GCACCGGAGACAATGGCTGCCGCTCAGTGCAGAACTTTCACAGACAATGCTACTAT	660	
DB	1078	GCACCGGAGACAATGGCTGCCGCTCAGTGCAGAACTTTCACAGACAATGCTACTAT	1137	
QY	661	TTTTCAAGTGAAGAAATTTTGGAGATGCAAGCTTTCTGTGAAGCAAGCTTCA	720	
DB	1138	TTTTCAAGTGAAGAAATTTTGGAGATGCAAGCTTTCTGTGAAGCAAGCTTCA	1197	
QY	721	CATCTTGTTCATTAACACTAGAGAGAAACAGCAATGATTAATAAAACAGATGTAGGG	780	
DB	1198	CATCTTGTTCATTAACACTAGAGAGAAACAGCAATGATTAATAAAACAGATGTAGGG	1257	
QY	781	AGAGAGAGCCATGATGCGCTTCACAGACTCAGAGCGTCAAAATGATGAAAGTGGCTG	840	
DB	1258	AGAGAGAGCCATGATGCGCTTCACAGACTCAGAGCGTCAAAATGATGAAAGTGGCTG	1317	
QY	841	GAGGGGATCTTCAGACTCAAAAATTTGAAAGCTGGAAGCGGATTAAGTGGGGCAT	900	
DB	1318	GAGGGGATCTTCAGACTCAAAAATTTGAAAGCTGGAAGCGGATTAAGTGGGGCAT	1377	
QY	901	GGCCATGAGGCGAGAGAACTGTGCTGAGTTATGCTGGGAGTGGAGCAATTC	960	
DB	1378	GGCCATGAGGCGAGAGAACTGTGCTGAGTTATGCTGGGAGTGGAGCAATTC	1437	
QY	961	CAATGGAAGAGCTCAATAATTCATTTGGGAAAAGACAGGAGACAGTACTGTATCT	1020	
DB	1438	CAATGGAAGAGCTCAATAATTCATTTGGGAAAAGACAGGAGACAGTACTGTATCT	1497	
QY	1021	GCATTA 1026		
DB	1498	GCATTA 1503		

RESULT 4				
AX490920	AX490920	2005 bp	DNA	linear PAT 16-AUG-2002
LOCUS	Sequence 27 from Patent WO020690.			
DEFINITION	AX490920			
ACCESSION	AX490920			
VERSION	AX490920.1	GI:22323797		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 Baker, K.P., Ferrara, N., Gerber, H., Gertlisen, M.E., Goddard, A., Godowski, P.J., Gunney, A.L., Hillan, K.J., Marsters, S.A., Pan, J., Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I. and Ye, W.			
TITLE	Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis			
JOURNAL	Patent: WO 020690-A 27 03-JAN-2002;			
Genentech, Inc. (US)				
FEATURES	Location/Qualifiers			
source	1..2005 /organism="Homo sapiens" /db_xref="taxon:9606"			
BASE COUNT	606 a 493 c 491 g 415 t			
ORIGIN				
Query Match	99.8%; Score 1024.4; DB 6; Length 2005;			
Best Local Similarity	99.9%; Pred. No. 8.7e-226;			
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1	ATGCACACAGATTGATGAGTGCAGGTTGAGACACTGAACTAGCAACATTCAGTGTAT	60	
DB	478	ATGCACACAGATTGATGAGTGCAGGTTGAGACACTGAACTAGCAACATTCAGTGTAT	537	
QY	61	ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACATA	120	
DB	538	ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAAATTTTACATA	597	
QY	121	CTTCAAGTTCACACGGGGCCCCAGGGTCCCAAGAGTGCAGAGAGTTCACAGGACCCCT	180	
DB	598	CTTCAAGTTCACACGGGGCCCCAGGGTCCCAAGAGTGCAGAGAGTTCACAGGACCCCT	657	
QY	181	GGCCCAACTGGCAACAAGGACAGAAAGAGAGAAGGGAGGAGCTGGACCACTGGCCCT	240	
DB	658	GGCCCAACTGGCAACAAGGACAGAAAGAGAGAAGGGAGGAGCTGGACCACTGGCCCT	717	
QY	241	GGGGGTGAGAGAGGCCCAATTGGACAGCTGTCCCGCCGAGAGCGTGGCGCAAGA	300	
DB	718	GGGGGTGAGAGAGGCCCAATTGGACAGCTGTCCCGCCGAGAGCGTGGCGCAAGA	777	
QY	301	TTCTAAGGCTCCAGGGCCCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGCCCTCAG	360	
DB	778	TTCTAAGGCTCCAGGGCCCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGCCCTCAG	837	
QY	361	GGCCCAAGTGGGAGACCCAGGCCCCCGGGCCCAACAGGACAGGAGGAGCTCCCGGCCCT	420	
DB	838	GGCCCAAGTGGGAGACCCAGGCCCCCGGGCCCAACAGGAGGAGGAGCTCCCGGCCCT	897	
QY	421	CAGGGCCCTCTGGCTTCCAGGAGCTTTCAGGAGCGGATTAAGTGGGGCAT	480	
DB	898	CAGGGCCCTCTGGCTTCCAGGAGCTTTCAGGAGCGGATTAAGTGGGGCAT	957	
QY	481	CTTCGGGGAGTGCAGGCTGCTGGGTTACAGGAGTGCAGAGGACCCCAAGGGCCCCC	540	
DB	958	CTTCGGGGAGTGCAGGCTGCTGGGTTACAGGAGTGCAGAGGACCCCAAGGGCCCCC	1017	
QY	541	GGCCCTCTGGGCCATCAGAGAGCGGTGTCGCCCTGGCCCTGCAGATGAGCCAAACCCG	600	
DB	1018	GGCCCTCTGGGCCATCAGAGAGCGGTGTCGCCCTGGCCCTGCAGATGAGCCAAACCCG	1077	

QY 601 GCACGGAGACATGCGTCCCGCCTCACTGAGAACTTCACAGACAAATGCTACTAT 660
|||||
Db 1078 GCACGGAGAGCAATAGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1137
QY 661 TTTTCAGTTGAGAAATAATTTTGTGAGATGCAAACTTTTCGTGAACAAGTCTTCA 720
|||||
Db 1138 TTTTCAGTTGAGAAATAATTTTGTGAGATGCAAACTTTTCGTGAACAAGTCTTCA 1197
QY 721 CATCTGTTTTCATAAACAAGTGAAGAAAGCAATGATATAAAACAGATGATGAGG 780
|||||
Db 1198 CATCTGTTTTCATAAACAAGTGAAGAAAGCAATGATATAAAACAGATGATGAGG 1257
QY 781 AGAGAGAGCCACTGATGCGCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 840
|||||
Db 1258 AGAGAGAGCCACTGATGCGCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 1317
QY 841 GATGGAGACATCTCCAGACTACAAAAATTTGAAAGCTGAGACGCGGATTAAGTGGGTCA 900
|||||
Db 1318 GATGGAGACATCTCCAGACTACAAAAATTTGAAAGCTGAGACGCGGATTAAGTGGGTCA 1377
QY 901 GGCCATGGGCGCAGAGAAAGTGTGCTGGGTGATTTATGCTGGGAGTGAAGCATTTTC 960
|||||
Db 1378 GGCCATGGGCGCAGAGAAAGTGTGCTGGGTGATTTATGCTGGGAGTGAAGCATTTTC 1437
QY 961 CATGTGAAGAGCTCAATTAATTTTCGAAAAAGACAGGAGAGACAGTACTGTATCT 1020
|||||
Db 1438 CATGTGAAGAGCTCAATTAATTTTCGAAAAAGACAGGAGAGACAGTACTGTATCT 1497
QY 1021 GCATTA 1026
|||||
Db 1498 GCATTA 1503

RESULT 5
AX047353
LOCUS AX047353 2641 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS Bandman, O., Hillman, J.L., Tang, Y.T., Lal, P., Yue, H., Baughn, M.R.,
Lu, D.A. and Azimzal, Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
Incyte Genomics, Inc. (US)

FEATURES
Source location/Qualifiers
1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN

Query Match 99.8%; Score 1024.4; DB 6; Length 2641;
Best Local Similarity 99.9%; Pred. No. 8.8e-226;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAACAAGATTGATGAGTGTGAGCTAGACATGAGTACCAACTTATCAGTAT 60
|||||
Db 912 ATGCAACAAGATTGATGAGTGTGAGCTAGACATGAGTACCAACTTATCAGTAT 971
QY 61 ATGGAAGAATGAGCTAGTACATCCAGCATGCTCAGCTCATCAAGATTTTACAATA 120
|||||
Db 972 ATGGAAGAATGAGCTAGTACATCCAGCATGCTCAGCTCATCAAGATTTTACAATA 1031
QY 121 CTACAAAGTCTCCACGGGCCCCAGGGGTCCAAAGAGTGACAGAGATCCAGGAGCCCT 180
|||||
Db 1032 CTACAAAGTCTCCACGGGCCCCAGGGGTCCAAAGAGTGACAGAGATCCAGGAGCCCT 1091

QY 181 GGCCCAACTGGCAACAAGGAGCAGAAAGAGAGAGAGGGGAGCCTGAGACCTGGCCCT 240
|||||
Db 1092 GGCCCAACTGGCAACAAGGAGCAGAAAGAGAGAGAGGGGAGCCTGAGACCTGGCCCT 1151
QY 241 GCGGGTGAGAGAGCCCAATTTGAGCAGCTGTCCCGCGAGAGCGTGGCGCAAGGA 300
|||||
Db 1152 GCGGGTGAGAGAGCCCAATTTGAGCAGCTGTCCCGCGAGAGCGTGGCGCAAGGA 1211
QY 301 TCTAAAGGCTCCCGAGGGCCCAAGGCTCCCGGTGTCCCTGGAAGCCCGGCCCTCAG 360
|||||
Db 1212 TCTAAAGGCTCCCGAGGGCCCAAGGCTCCCGGTGTCCCTGGAAGCCCGGCCCTCAG 1271
QY 361 GGCCCGCAGTGGGAGCCAGGCCCCCGGGGCCACAGGCAAAAGAGGACTCCCGGCCCT 420
|||||
Db 1272 GGCCCGCAGTGGGAGCCAGGCCCCCGGGGCCACAGGCAAAAGAGGACTCCCGGCCCT 1331
QY 421 CAGGGCCCTCCTGGCTTCACAGGAGCTTCAGGGCACCCTGGGGAGCCTGGGGTGCCTGA 480
|||||
Db 1332 CAGGGCCCTCCTGGCTTCACAGGAGCTTCAGGGCACCCTGGGGAGCCTGGGGTGCCTGA 1391
QY 481 CCTCGGGGAGCTGCCAGGCTTGCCTGGGGTACAGGCTATGCCAGGCCCCCAAGGCCCTCC 540
|||||
Db 1392 CCTCGGGGAGCTGCCAGGCTTGCCTGGGGTACAGGCTATGCCAGGCCCCCAAGGCCCTCC 1451
QY 541 GGCCCTCCTGGGCGCATCAGGAGGCGGTGGTCCCTGGGCCCTGCAGATGAGCCACGCCG 600
|||||
Db 1452 GGCCCTCCTGGGCGCATCAGGAGGCGGTGGTCCCTGGGCCCTGCAGATGAGCCACGCCG 1511
QY 601 GCACCGGAGGACAATGGCTGCCCGCTCACTGAGAAACTTTCACAGACAAATGCTACTAT 660
|||||
Db 1512 GCACCGGAGGACAATGGCTGCCCGCTCACTGAGAAACTTTCACAGACAAATGCTACTAT 1571
QY 661 TTTTCAGTTGAGAAATAATTTTGTGAGATGCAAGCTTTTCTGTGAACAAGTCTTCA 720
|||||
Db 1572 TTTTCAGTTGAGAAATAATTTTGTGAGATGCAAGCTTTTCTGTGAACAAGTCTTCA 1631
QY 721 CATCTGTTTTCATAAACAAGTGAAGAAAGCAACATGATATAAAACAGATGATGAGG 780
|||||
Db 1632 CATCTGTTTTCATAAACAAGTGAAGAAAGCAACATGATATAAAACAGATGATGAGG 1691
QY 781 AGAGAGAGCCACTGATGCGCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 840
|||||
Db 1692 AGAGAGAGCCACTGATGCGCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 1751
QY 841 GATGGAGACATCTCCAGACTACAAAAATTTGAAAGCTGAGACGCGGATTAAGTGGGTCA 900
|||||
Db 1752 GATGGAGACATCTCCAGACTACAAAAATTTGAAAGCTGAGACGCGGATTAAGTGGGTCA 1811
QY 901 GGCCATGGGCGCAGAGAAAGTGTGCTGGGTGATTTATGCTGGGAGTGAAGCATTTTC 960
|||||
Db 1812 GGCCATGGGCGCAGAGAAAGTGTGCTGGGTGATTTATGCTGGGAGTGAAGCATTTTC 1871
QY 961 CAATGTGAAGAGCTCAATTAATTTTCGAAAAAGACAGGAGAGACAGTACTGTATCT 1020
|||||
Db 1872 CAATGTGAAGAGCTCAATTAATTTTCGAAAAAGACAGGAGAGACAGTACTGTATCT 1931
QY 1021 GCATTA 1026
|||||
Db 1932 GCATTA 1937

RESULT 6
AB038518
LOCUS AB038518 3058 bp mRNA linear PRI 08-MAR-2001
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type I, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE Homo sapiens tissue_lhb:Placenta cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible]

```

/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDFAEEBEVPSFGFRFGIHEGTQCKINNMALKEFISVILY
ILCALITVIALIGYVVEKMDNVSDEGTSKQTDLTAVESDLKLGQDQKAL
STNSELSTFRSDIDLROOLETEKTSKNDLTLEKLANGSLDROSLKETLONN
SFLITTVNKTLOAYNGYTNLQODNTVNOGNQSQMSQSVYIMNLINLTVOQOR
LISNLQOSVDTSLAIORIKNDPQNLQOVFLQAKDQDWLKEKVSLOTLANNSALA
KANNDTLEDMSQLSFTGQMDNITTTISQANDQSLKLDLKHQDENTRAVFSOLEE
RPOVETDVIWNISISYTAHRLRTLSLNDVWTCTDITLRTDLDLSLNTLVNI
RDSISLRQODMRSKLDTEEVANLSVVEEMKLVDSKGLIKNTTILQGPGRGP
KDRGSQGPPTGPKNGKQGEKPEPGEKRGITGVGPEVDSKSGSGSPK
GSRGSPKRGPGSPGSDRGPPEKDGIRGQGPGRGLOGTGEBVPEPRGIRP
LPGVPMGPPEKPPPEPSPGMEPLQDNETPASEVNGCPPEHMKNTFDKCTYFSL
KEILDALKEEDKSSHLVINSREQDQWIKHTVGRSHVIGLADSEQSEMKWLDG
SPVDYKMKMKAQPDNMWGSQHGPEDCAGLIVAGWNDFQCDINNFICKEFEAVPSS
IL"
polyA_signal 3246.3251
/gene="srcl"
BASE COUNT 1013 a 757 c 759 g 762 t
ORIGIN
Query Match 78.0%; Score 800.4; DB 10; Length 3291;
Best Local Similarity 86.3%; Pred No. 4.2e-174;
Matches 885; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1 ATGCAACAAGATTGATGAGTGGAGTTAGACATGAAGTACCAATTACATGATT 60
DB 1277 ATGCAACAAGATTGATGAGTGGAGTTAGACATGAAGTACCAATTACATGATT 1336
QY 61 ATGGAAGAATGAGTACTAGTACTCCAAAGCATGTCAGTCTATCAAGATTTTACAATA 120
DB 1337 ATGGAAGAATGAGTACTAGTACTCCAAAGCATGTCAGTCTATCAAGATTTTACAATA 1396
QY 121 CTACAGAGTCCACCGGGCCCAAGGGTCCAAAGAGTGTACAGAGATCCAGAGACCCCT 180
DB 1397 CTACAGAGTCCCTCTGCTGGCCCAAGAGTCCAAAGAGTGTACAGAGATCTCAGAGACCACT 1456
QY 181 GGGCCACTGGCAACAAGGAGACAGAAAGAGAGAGAGGAGGAGGAGCCTTGACCACTGGCCCT 240
DB 1457 GGTCAACTGGCAACAAGGAGACAGAAAGAGAGAGAGAGGAGGAGCCTTGACCACTGGCCCT 1516
QY 241 GGGGATGAGAGAGGCCAATTGGACAGTGTCCCGGGAGAGCGTGGCGGCAAAAGGA 300
DB 1517 GGGGATGAGAGAGGCCAATTGGACAGTGTCCCGGGAGAGCGTGGCGGCAAAAGGA 1576
QY 301 TCTAAAGGCTCCAGGGGCCCAAAAGGCTCCCGGTTCCCTGGGAAGGCCGGCCCTCAG 360
DB 1577 TCTAAAGGCTCCAGGGGCCCAAAAGGATCTCTGGGTTCCCGGAGGAGACCTTGCCCTCA 1636
QY 361 GGGCCAGTGGGAGACCCAGGCCCCCGGGGCCCAACAGGAGAGAGGAGACTCCCGGGCCCT 420
DB 1637 GGCACATAGTGGGAGACCCAGGACCAGGTCACAGGCAAGGATGACTCCCTTGCCCT 1696
QY 421 CAGGGCCCTCTGCTTCCAGGAGCTTCAAGGAGTGGAGAGCGCTGGGGTGGCTGGGA 480
DB 1697 CAGGGCCCTCTGCTTCCAGGAGCTTCAAGGAGTGGAGAGCGCTGGAGTACCTGGA 1756
QY 481 CCTCGGGAGTGGCAGGCTTGGCTGGGATACAGGATGCGAAGGCCCAAGGGCCCCCCC 540
DB 1757 CCTCGGGAGTGGCAGGCTTGGCAGGAGTGGCAGGAGTGGCTTGAAGGAGCAACCT 1816
QY 541 GGGCCCTCTGCGCCATCAGAGAGGAGTGGTCCCTGGCCCTGAGAAATGAGCCAAACCCG 600
DB 1817 GGGCCCTCTGCGCCCTCAGAGAGAAATGAGGAGCTTGGCTTGAAGTACCAACCCCA 1876
QY 601 GACACGGAGAGAAATGGCTGCCGCTCACTGGAAGAACTTCAAGAGAAATGCTACTAT 660
DB 1877 GCATCAGAGAGTCAACGAGTCCGCTCACTGGAAGAACTTCAAGATTAATGCTACTAT 1936
QY 661 TTTTTCAGTGAAGAAGAAATTTTGGATGCAAAAGCTTTGCTGTAAGACAAAGCTTCA 720
DB 1937 TTTTTCAGTGAAGAAGAAATTTTGGATGCAAAAGCTTTTCTGTGAAGACAAATCTTCC 1996

```

```

QY 721 CATCTGTTTTCATTAACACTAGAGAGAACAGCAATGATAAAAAACAGATGTAGGG 780
DB 1997 CATCTGTTTTCATTAACACTAGAGAGAACAGCAATGATAAAAAACAGATGTAGGG 2056
QY 781 AGAGAGAGCCACGATGTGGCCCTCACAGACTCAGACCGTCAAAATGAATGAGTGGCTG 840
DB 2057 AGAGAGAGCCACGATGTGGCCCTCACAGACTCAGAACAGAGGCAATGAGTGGCTG 2116
QY 841 GAAGGACATCTCCACACTCAAAAAATTTGGAAGCTGAGACGGCGATTAATGGGGTCAAT 900
DB 2117 GAGGGGTCACCTTTGATTACAAAACTGGAAGCTGAGAACACAGATTAATGGGGCAGT 2176
QY 901 GGCATGAGGCCAGAGAAAGACTGTGCTGGTGTGATTTATGCTGGGAGTGAACGATTTTC 960
DB 2177 GGCATGAGGCCAGAGAAAGACTGTGCTGGTGTGATTTATGACAGAGAGTGAATGACTTC 2236
QY 961 CATGTGAGAGCTCAATATCAATTCATTGGGAAAAAGACAGGAGACAGTATGATCT 1020
DB 2237 CATGTGATGAATCAATATCAATTCATTGTGAGAGGAAAGGAGGACAGTACCATCATCC 2296
QY 1021 GCATTA 1026
DB 2297 ATATTA 2302
QY 1021 GCATTA 1026
DB 2297 ATATTA 2302
RESULT 9
AB052103 4330 bp mRNA linear PRI 08-MAR-2001
LOCUS AB052103
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS
SOURCE
ORGANISM Homo sapiens
CDNA to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Nakamura,K., Funkhoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
REFERENCE 2 (bases 1 to 4330)
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
School of Medicine, Division of Biochemistry, Biomedical Research
Center, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@bich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)
FEATURES
source
location/Qualifiers
1..4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..4330
/gene="SRCL"
3..1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:13365553"
/translation="MKDFAEEBEVPSFGFRFGIHEGTQCKINNMALKEFISVILY
ILCALITVIALIGYVVEKMDNVSDEGTSKQTDLTAVESDLKLGQDQKAL
STNSELSTFRSDIDLROOLETEKTSKNDLTLEKLANGSLDROSLKETLONN
SFLITTVNKTLOAYNGYTNLQODNTVNOGNQSQMSQSVYIMNLINLTVOQOR
LISNLQOSVDTSLAIORIKNDPQNLQOVFLQAKDQDWLKEKVSLOTLANNSALA
KANNDTLEDMSQLSFTGQMDNITTTISQANDQSLKLDLKHQDENTRAVFSOLEE
RPOVETDVIWNISISYTAHRLRTLSLNDVWTCTDITLRTDLDLSLNTLVNI
RDSISLRQODMRSKLDTEEVANLSVVEEMKLVDSKGLIKNTTILQGPGRGP
KDRGSQGPPTGPKNGKQGEKPEPGEKRGITGVGPEVDSKSGSGSPK
GSRGSPKRGPGSPGSDRGPPEKDGIRGQGPGRGLOGTGEBVPEPRGIRP
LPGVPMGPPEKPPPEPSPGMEPLQDNETPASEVNGCPPEHMKNTFDKCTYFSL
KEILDALKEEDKSSHLVINSREQDQWIKHTVGRSHVIGLADSEQSEMKWLDG
SPVDYKMKMKAQPDNMWGSQHGPEDCAGLIVAGWNDFQCDINNFICKEFEAVPSS
IL"

```



```

clone_end:SP6
vector_side:left"
misc_feature
18044..22424
/note="assembly_fragment"
misc_feature
22525..38094
/note="assembly_fragment"
misc_feature
38195..49220
/note="assembly_fragment"
misc_feature
49321..66394
/note="assembly_fragment"
misc_feature
66495..91692
/note="assembly_fragment"
misc_feature
91793..120869
/note="assembly_fragment"
misc_feature
120970..169088
/note="assembly_fragment"
misc_feature
clone_end:T7
vector_side:right"
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN
```

Query Match	47.6%	Score 488.8	DB 2	Length 169088
Best Local Similarity	99.6%	Pred. No. 4e-102		
Matches 490	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY 126	AGGTCACACGGGGCCCCAGAGGTCCAAAGAGTGCACAGAGATCCAGGGACCCCTGTGGCCC	185		
Db 153817	AGGTCACACGGGGCCCCAGAGGTCCAAAGAGTGCACAGAGATCCAGGGACCCCTGTGGCCC	153758		
QY 186	AACGTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGGCTTGAGCCACTGTGGCCCTGTGGGG	245		
Db 153757	AACGTGGCAACAAGGAGACAGAAAGAGAGAGAGGGGAGGCTTGAGCCACTGTGGCCCTGTGGGG	153698		
QY 246	TGAGAGAGGGCCCAATTGAGACAGCTGTGCCCCCGGAGAGAGGTGTGGCGCAAAAGATCTTAA	305		
Db 153697	TGAGAGAGGGCCCAATTGAGACAGCTGTGCCCCCGGAGAGAGGTGTGGCGCAAAAGATCTTAA	153638		
QY 306	AGGCTCCAGAGGGCCCAAGAGCTCCCGTGTTCCTCGTGGAAAGCCCGGCTCCAGAGGCC	365		
Db 153637	AGGCTCCAGAGGGCCCAAGAGCTCCCGTGTTCCTCGTGGAAAGCCCGGCTCCAGAGGCC	153578		
QY 366	CAGTGGGGACCCAGAGCCCCCGGGCCACACAGGCAAGAGAGGACTCCCGGCTTCAGGG	425		
Db 153577	CAGTGGGGACCCAGAGCCCCCGGGCCACACAGGCAAGAGAGGACTCCCGGCTTCAGGG	153518		
QY 426	CCCTTCCTGTGCTTCCAGAGGACTTCAGGGACCGCTTGTGGAGCCTGGGGTGTCTGTGACCTCG	485		
Db 153517	CCCTTCCTGTGCTTCCAGAGGACTTCAGGGACCGCTTGTGGAGCCTGGGGTGTCTGTGACCTCG	153458		
QY 486	GGGACTGCGCAGGCGTTCCTGGGGTACCAAGGCATGCGCAGGCCCCCAAGGGCCCCCGGGCCC	545		
Db 153457	GGGACTGCGCAGGCGTTCCTGGGGTACCAAGGCATGCGCAGGCCCCCAAGGGCCCCCGGGCCC	153398		
QY 546	TCCGTGGCCATTCAGAGCGGTGTGTCGCCCTGGGCCCTGGAGATGAGCAACCCCGGACCC	605		
Db 153397	TCCGTGGCCATTCAGAGCGGTGTGTCGCCCTGGGCCCTGGAGATGAGCAACCCCGGACCC	153338		
QY 606	GGAGGACACATGG 617			
Db 153337	GGAGGACACATGG 153326			
RESULT 11				
AP000915/c				
LOCUS	AP000915	188439 bp	DNA	linear
DEFINITION	Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete			
ACCESSION	AP000915			
VERSION	AP000915.5	GI:20334314		
KEYWORDS	HTG.			
SOURCE	Homo sapiens DNA, clone:RP11-720L2.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

REFERENCE	1	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS		Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE		Homo sapiens genomic DNA
JOURNAL		Published only in Database (1999)
REFERENCE		2 (bases 1 to 188439)
AUTHORS		Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE		Direct Submission
JOURNAL		Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: hattori@gsc.riken.go.jp, URL: http://ngp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT		On Apr 26, 2002 this sequence version replaced g1:9188470.
FEATURES		Location/Qualifiers
SOURCE		1. 188439
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="18"
		/map="18p"
		/clone="RP11-720L2"
BASE COUNT		54003 a 41134 c 40027 g 53275 t
ORIGIN		

Query Match	47.6%	Score 488.8	DB 9	Length 188439
Best Local Similarity	99.6%	Pred. No. 4e-102		
Matches 490	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY 126	AGGTCCACCGGGGGCCCCAGGGGCTCCAAAGAGTGACAGAGAGATCCCAAGGGACCCCTTGCGCC	185		
Db 44938	AGGTCCACCGGGGGCCCCAGGGGCTCCAAAGAGTGACAGAGAGATCCCAAGGGACCCCTTGCGCC	44879		
QY 186	AACGTGCAACCAAGGAGACAGAAAGAGAGAGAGGGGAGCGCTGACACACTGGCCCGCGGG	245		
Db 44878	AACGTGCAACCAAGGAGACAGAAAGAGAGAGAGGGGAGCGCTGACACACTGGCCCGCGGG	44819		
QY 246	TGAGAGAGGCCCAATTGGACACAGCTGTGCCCCCGAGAGCGTGGCGCAAAAGATCTAA	305		
Db 44818	TGAGAGAGGCCCAATTGGACACAGCTGTGCCCCCGAGAGCGTGGCGCAAAAGATCTAA	44759		
QY 306	AGGCTCCAGGGGGCCCCAAAGGCTCCCGGTGTCCCTTGAGAAAGCCCGGCTTCAGGGGCC	365		
Db 44758	AGGCTCCAGGGGGCCCCAAAGGCTCCCGGTGTCCCTTGAGAAAGCCCGGCTTCAGGGGCTC	44699		
QY 366	CAGTGGGGAGCCACAGGCCCCCGGGGCCACACAGGCAAAAGAGAGACTCCCGGCTTCAGGG	425		
Db 44698	CAGTGGGGAGCCACAGGCCCCCGGGGCCACACAGGCAAAAGAGAGACTCCCGGCTTCAGGG	44639		
QY 426	CCCTTCCTGCTTCCAGGAGACTTCAAGGGCACCGTTGGGGAGGCTTGGGGTGCCTGGACCTCG	485		
Db 44638	CCCTTCCTGCTTCCAGGAGACTTCAAGGGCACCGTTGGGGAGGCTTGGGGTGCCTGGACCTCG	44579		
QY 486	GGGAGCTGCAGAGCTTGGCTGGGGTACACAGAGCATCCACAGGCCCAAGGGGCCCGCGGGCC	545		
Db 44578	GGGAGCTGCAGAGCTTGGCTGGGGTACACAGAGCATCCACAGGCCCAAGGGGCCCGCGGGCC	44519		
QY 546	TCTGGGCCCATCAGAGCGGTGTGTCCTTGCGCTTGCGCTGCAGATAGCCACCCCGGACAC	605		
Db 44518	TCTGGGCCCATCAGAGCGGTGTGTCCTTGCGCTTGCGCTGCAGATAGCCACCCCGGACAC	44459		
QY 606	GGAGGACAAATGG 617			
Db 44458	GGAGGACAAATGG 44447			
RESULT 12				
LOCUS AC024368	71044 bp	DNA	linear	HTG 28-FEB-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-179k3 map 11, LOW-PASS				
ACCESSION AC024368				

```

VERSION      AC024368.1 GI:7108157
KEYWORDS     HTG: HTGS_PHASE0.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 71044)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 11, clone RP11-179K3
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 71044)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
              Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
              Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
              Fennestor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
              Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
              Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
              Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
              Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
              McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
              Menous,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
              Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M.,
              Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
              Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
              Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Subramanian,A., Talamas,J., Testfaye,S., Theodore,J., Tirelli,A.,
              Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
              Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zimmer,A. and
              Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: L6145
              Center clone name: 179_K_3
              -----
              * NOTE: This record contains 90 individual
              * sequencing reads that have not been assembled into
              * contigs. Runs of N are used to separate the reads
              * and the order in which they appear is completely
              * arbitrary. Low-pass sequence sampling is useful for
              * identifying clones that may be gene-rich and allows
              * overlap relationships among clones to be deduced.
              * However, it should not be assumed that this clone
              * will be sequenced to completion. In the event that
              * the record is updated, the accession number will
              * be preserved.
              *
              * 1
              * 699 798: contig of 698 bp in length
              * 799 1495: contig of 697 bp in length
              * 1496 1595: gap of 100 bp
              * 1596 2284: contig of 689 bp in length
              * 2285 2384: gap of 100 bp
              * 2385 3056: contig of 672 bp in length
              * 3057 3156: gap of 100 bp
              * 3157 3851: contig of 695 bp in length
              * 3852 3951: gap of 100 bp
              * 3952 4627: contig of 676 bp in length
              * 4628 4727: gap of 100 bp
              * 4728 5409: contig of 682 bp in length
              * 5410 5509: gap of 100 bp
              * 5510 6183: contig of 674 bp in length
              -----
              * 6184 6283: gap of 100 bp
              * 6284 6984: contig of 701 bp in length
              * 6985 7084: gap of 100 bp
              * 7085 7768: contig of 684 bp in length
              * 7769 7868: gap of 100 bp
              * 7869 8550: contig of 682 bp in length
              * 8551 8650: gap of 100 bp
              * 8651 9352: contig of 682 bp in length
              * 9353 9452: gap of 100 bp
              * 9453 10140: contig of 688 bp in length
              * 10141 10240: gap of 100 bp
              * 10241 10922: contig of 682 bp in length
              * 10923 11022: gap of 100 bp
              * 11023 11709: contig of 687 bp in length
              * 11710 11809: gap of 100 bp
              * 11810 12511: contig of 702 bp in length
              * 12512 12611: gap of 100 bp
              * 12612 13504: contig of 693 bp in length
              * 13505 13404: gap of 100 bp
              * 13405 14095: contig of 691 bp in length
              * 14096 14195: gap of 100 bp
              * 14196 14881: contig of 686 bp in length
              * 14882 14981: gap of 100 bp
              * 14982 15674: contig of 693 bp in length
              * 15675 15774: gap of 100 bp
              * 15775 16460: contig of 686 bp in length
              * 16461 16560: gap of 100 bp
              * 16561 17250: contig of 690 bp in length
              * 17251 17350: gap of 100 bp
              * 17351 18040: contig of 690 bp in length
              * 18041 18140: gap of 100 bp
              * 18141 18836: contig of 696 bp in length
              * 18837 18936: gap of 100 bp
              * 18937 19621: contig of 686 bp in length
              * 19622 19721: gap of 100 bp
              * 19722 20425: contig of 704 bp in length
              * 20426 20525: gap of 100 bp
              * 20526 21224: contig of 699 bp in length
              * 21225 21324: gap of 100 bp
              * 21325 22017: contig of 693 bp in length
              * 22018 22117: gap of 100 bp
              * 22118 22815: contig of 696 bp in length
              * 22816 22915: gap of 100 bp
              * 22916 23609: contig of 694 bp in length
              * 23610 23709: gap of 100 bp
              * 23710 24402: contig of 693 bp in length
              * 24403 24502: gap of 100 bp
              * 24503 25191: contig of 689 bp in length
              * 25192 25291: gap of 100 bp
              * 25292 25981: contig of 690 bp in length
              * 25982 26081: gap of 100 bp
              * 26082 26774: contig of 693 bp in length
              * 26775 26874: gap of 100 bp
              * 26875 27563: contig of 688 bp in length
              * 27564 27663: gap of 100 bp
              * 27664 28355: contig of 692 bp in length
              * 28356 28455: gap of 100 bp
              * 28456 29145: contig of 690 bp in length
              * 29146 29245: gap of 100 bp
              * 29246 29943: contig of 696 bp in length
              * 29944 30043: gap of 100 bp
              * 30044 30745: contig of 702 bp in length
              * 30746 30845: gap of 100 bp
              * 30846 31535: contig of 690 bp in length
              * 31536 31635: gap of 100 bp
              * 31636 32329: contig of 694 bp in length
              * 32330 32429: gap of 100 bp
              * 32430 33114: contig of 685 bp in length
              * 33115 33214: gap of 100 bp
              * 33215 33898: contig of 684 bp in length
              * 33899 33998: gap of 100 bp
              * 33999 34694: contig of 696 bp in length
              * 34695 34794: gap of 100 bp

```


21433	42072	contig	of	20640	bp	in	length
42173	54336	contig	of	12164	bp	in	length
54337	66766	contig	of	12330	bp	in	length
66867	77849	contig	of	10983	bp	in	length
77950	87798	contig	of	9849	bp	in	length
87899	99510	contig	of	11612	bp	in	length
99611	107273	contig	of	7663	bp	in	length
107374	113189	contig	of	5816	bp	in	length
113290	120559	contig	of	7270	bp	in	length
120660	126468	contig	of	5809	bp	in	length
126569	131619	contig	of	5051	bp	in	length
131720	136810	contig	of	5091	bp	in	length
136911	139925	contig	of	3015	bp	in	length
136911	139925	contig	of	4570	bp	in	length
140026	144332	contig	of	4207	bp	in	length
144332	149003	contig	of	3361	bp	in	length
149003	152918	contig	of	3316	bp	in	length
153019	156399	contig	of	3916	bp	in	length
156500	158933	contig	of	2434	bp	in	length
159034	161884	contig	of	2851	bp	in	length
161985	163301	contig	of	2317	bp	in	length
164402	166300	contig	of	1889	bp	in	length
166401	166580	contig	of	2180	bp	in	length
166861	171270	contig	of	2590	bp	in	length
171371	174092	contig	of	2722	bp	in	length
174193	176093	contig	of	1903	bp	in	length
176196	177890	contig	of	1695	bp	in	length
177991	179786	contig	of	1796	bp	in	length
179887	181478	contig	of	1592	bp	in	length
181579	182895	contig	of	1317	bp	in	length
182996	184841	contig	of	1846	bp	in	length
184942	186345	contig	of	1404	bp	in	length
186446	187635	contig	of	1190	bp	in	length
Sequence updated (28-May-2000).							
NOTE: This is a 'working draft' sequence. It currently							
consists of 33 contigs. The true order of the pieces							
is not known and their order in this sequence record is							
arbitrary. Gaps between the contigs are represented as							
runs of N, but the exact sizes of the gaps are unknown.							
This record will be updated with the finished sequence							
as soon as it is available and the accession number will							
be preserved.							
1	21332:	contig	of	21332	bp	in	length
21333	21432:	gap	of	100	bp		
21433	42072:	contig	of	20640	bp	in	length
42073	42172:	gap	of	100	bp		
42173	54336:	contig	of	12164	bp	in	length
54337	54436:	gap	of	100	bp		
54437	66766:	contig	of	12330	bp	in	length
66767	66866:	gap	of	100	bp		
66867	77849:	contig	of	10983	bp	in	length
77850	77949:	gap	of	100	bp		
77950	87798:	contig	of	9849	bp	in	length
87799	87898:	gap	of	100	bp		
87899	99510:	contig	of	11612	bp	in	length
99511	99610:	gap	of	100	bp		
99611	107273:	contig	of	7663	bp	in	length
107274	107373:	gap	of	100	bp		
107374	113189:	contig	of	5816	bp	in	length
113190	113289:	gap	of	100	bp		
113290	113289:	contig	of	7270			

	*	149003	152918:	contig of 3916 bp	in length
	*	152919	153018:	gap of 100 bp	
	*	153019	156399:	contig of 3381 bp	in length
	*	156400	156499:	gap of 100 bp	
	*	156500	158933:	contig of 2434 bp	in length
	*	158934	159033:	gap of 100 bp	
	*	159034	161884:	contig of 2851 bp	in length
	*	161885	161984:	gap of 100 bp	
	*	161985	164301:	contig of 2317 bp	in length
	*	164302	164401:	gap of 100 bp	
	*	164402	166300:	contig of 1899 bp	in length
	*	166301	166400:	gap of 100 bp	
	*	166401	168580:	contig of 2180 bp	in length
	*	168581	168680:	gap of 100 bp	
	*	168681	171270:	contig of 2590 bp	in length
	*	171271	171370:	gap of 100 bp	
	*	171371	174092:	contig of 2722 bp	in length
	*	174093	174192:	gap of 100 bp	
	*	174193	176095:	contig of 1903 bp	in length
	*	176096	176195:	gap of 100 bp	
	*	176196	177890:	contig of 1695 bp	in length
	*	177891	177990:	gap of 100 bp	
	*	177991	179786:	contig of 1796 bp	in length
	*	179787	179886:	gap of 100 bp	
	*	179887	181478:	contig of 1592 bp	in length
	*	181479	181578:	gap of 100 bp	
	*	181579	182895:	contig of 1317 bp	in length
	*	182896	182995:	gap of 100 bp	
	*	182996	184841:	contig of 1846 bp	in length
	*	184842	184941:	gap of 100 bp	
	*	184942	186345:	contig of 1404 bp	in length
	*	186346	186445:	gap of 100 bp	
	*	186446	187635:	contig of 1190 bp	in length.
FEATURES					
source					
			1.	187635	
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/chromosome="18"		
			/map="18p11.3"		
			/clone="RP11-815L4"		
			1.	21332	
misc_feature			/note="assembly_fragment"		
			21433.	42072	
misc_feature			/note="assembly_fragment"		
			42173.	54336	
misc_feature			/note="assembly_fragment"		
			54437.	66766	
misc_feature			/note="assembly_fragment"		
			66867.	77849	
misc_feature			/note="assembly_fragment"		
			77950.	87798	
misc_feature			/note="assembly_fragment_clone_end:r7 vector_side:right"		
			87899.	99510	
misc_feature			/note="assembly_fragment"		
			99611.	107273	
misc_feature			/note="assembly_fragment"		
			107374.	113189	
misc_feature			/note="assembly_fragment_clone_end:SP6 vector_side:left"		
			113290.	120559	
misc_feature			/note="assembly_fragment"		
			120660.	126468	
misc_feature			/note="assembly_fragment"		
			126569.	131619	
misc_feature			/note="assembly_fragment"		
			131720.	136810	
misc_feature			/note="assembly_fragment"		
			136911.	139925	
misc_feature			/note="assembly_fragment"		
			140026.	144232	
misc_feature			/note="assembly_fragment"		
			144333.	148902	
misc_feature			/note="assembly_fragment"		
			149003.	152918	

```
misc_feature      /note="assembly-fragment"
                  153019..156399

Query Match      45.4%; Score 466.2; DB 2; Length 187635;
Best Local Similarity 96.3%; Pred. No.6.5e-97;
Matches 488; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 111 TTTTACATATCTACAAAGTCCACCGGGCCAGGGGTCCAAAGAGTGACAGAGATGCCA 170
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68312 TTCTTTATGCTTTAGTGTCCACCGGGCCAGGGGTCCAAAGAGTGACAGAGATGCCA 68371
QY 171 GGGACCCCTCGGCTCCAACTGGCAACAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAG 230
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68372 GGGACCCCTCGGCTCCAACTGGCAACAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAG 68431
QY 231 ACCTGGCCCTCGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68432 ACCTGGCCCTCGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 68491
QY 291 CGGCAAGAGATCTAAAGGCTCCAGAGGCCCCAAAGGCTCCGTGGTCCCTGGGAAGCC 350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68492 CGGCAAGAGATCTAAAGGCTCCAGAGGCCCCAAAGGCTCCGTGGTCCCTGGGAAGCC 68551
QY 351 CGGCTCCAGAGGCCCCAGAGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGAGACT 410
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68552 CGGCTCCAGAGGCCCCAGAGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGAGACT 68610
QY 411 CCCCAGGCTCCAGAGGCTCCGTGGTCCAGAGACTTCAGAGGACCGTTGGAGAGCTGG 470
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68611 CCCCAGGCTCCAGAGGCTCCGTGGTCCAGAGACTTCAGAGGACCGTTGGAGAGCTGG 68670
QY 471 GGTGCTCGAGCTCGGAGGACTCCAGAGCTTCCTGGGGTACAGAGCATGCCAGAGCCCCAA 530
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68671 GGTGCTCGAGCTCGGAGGACTCCAGAGCTTCCTGGGGTACAGAGCATGCCAGAGCCCCAA 68730
QY 531 GGGCCCCCGGGCTCCGTGGCCATCAGAGAGGGGTGGTCCCTGGGCTCGAGAGATGA 590
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68731 GGGCCCCCGGGCTCCGTGGCCATCAGAGAGGGGTGGTCCCTGGGCTCGAGAGATGA 68790
QY 591 GCCAACCCCGGACCGAGAGACATGG 617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68791 GCCAACCCCGGACCGAGAGACATGG 68817

RESULT 14
LOCUS      AP000900      178022 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKING
ACCESSION  AP000900
VERSION    AP000900.3 GI:8119043
KEYWORDS   HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE     Homo sapiens DNA, clone:RP11-683J11.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 178022)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
            Homo sapiens 178,022 genomic DNA of 18p11.3
            Published Only in Database (1999)
            2 (bases 1 to 178022)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
            Direct Submission
            Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Katsato Univ., 1-15-1 Katsato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
            On May 31, 2000 this sequence version replaced gi:6997403.
COMMENT    ----- Genome Center
```

```
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information -----
Center project name: HumDrafl18
Center clone name: RP11-683J11
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-merman; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161944 bases at least Q40
Consensus quality: 170357 bases at least Q30
Consensus quality: 174322 bases at least Q20
Insert size: 176122; sum-of-ctrls
Quality coverage: 4.56x in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently consists of
20 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1
24384 24283 contig of 24283 bp in length
46976 46875 contig of 22492 bp in length
65159 18083 contig of 18083 bp in length
81423 81322 contig of 16164 bp in length
98924 98823 contig of 17401 bp in length
110169 110068 contig of 11145 bp in length
122817 122716 contig of 12548 bp in length
128817 133089 contig of 10273 bp in length
141307 141206 contig of 8017 bp in length
149151 156105 contig of 6955 bp in length
156206 162048 contig of 5843 bp in length
162149 165118 contig of 2970 bp in length
165219 167480 contig of 2262 bp in length
167581 169986 contig of 2406 bp in length
170087 171909 contig of 1823 bp in length
172010 172317 contig of 308 bp in length
172418 174653 contig of 2246 bp in length
174764 176271 contig of 1508 bp in length
176372 178022 contig of 1651 bp in length

Sequence updated (24-Dec-1999).
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
24284 24283 contig of 24283 bp in length
24384 24383 contig of 100 bp
46876 46875 contig of 22492 bp in length
46976 46975 contig of 100 bp
65059 65158 contig of 18083 bp in length
65159 65158 contig of 100 bp
81323 81422 contig of 16164 bp in length
81423 81422 contig of 100 bp
98824 98923 contig of 17401 bp in length
98924 98923 contig of 100 bp
110069 110168 contig of 11145 bp in length
110169 110168 contig of 100 bp
122717 122716 contig of 12548 bp in length
122717 122816 contig of 100 bp
128817 133089 contig of 10273 bp in length
133090 133189 contig of 100 bp
133190 141206 contig of 8017 bp in length
141207 141306 contig of 100 bp
141307 149050 contig of 7744 bp in length
```

```

* 149051 149150: gap of 100 bp
* 149151 156105: contig of 6955 bp in length
* 156106 156205: gap of 100 bp
* 156206 162048: contig of 5843 bp in length
* 162049 162148: gap of 100 bp
* 162149 165118: contig of 2970 bp in length
* 165119 165218: gap of 100 bp
* 165219 167480: contig of 2262 bp in length
* 167481 167580: gap of 100 bp
* 167581 169986: contig of 2406 bp in length
* 169987 170086: gap of 100 bp
* 170087 171909: contig of 1823 bp in length
* 171910 172009: gap of 100 bp
* 172010 172317: contig of 308 bp in length
* 172318 172417: gap of 100 bp
* 172418 174663: contig of 2246 bp in length
* 174664 174763: gap of 100 bp
* 174764 176271: contig of 1508 bp in length
* 176272 176371: gap of 100 bp
* 176372 178022: contig of 1651 bp in length.

```

FEATURES

source

```

1. 178022
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="18"
   /map="18p11.3"
   /clone="RP11-683J11"
1. 24283
   /note="assembly-fragment"
misc_feature
24384..46875
   /note="assembly-fragment"
misc_feature
46976..65058
   /note="assembly-fragment"
misc_feature
65159..81322
   /note="assembly-fragment"
misc_feature
81423..98623
   /note="assembly-fragment"
misc_feature
98924..110068
   /note="assembly-fragment"
misc_feature
110169..122716
   /note="assembly-fragment"
misc_feature
122817..133069
   /note="assembly-fragment"
misc_feature
133190..141206
   /note="assembly-fragment"
misc_feature
141307..149050
   /note="assembly-fragment"
misc_feature
149151..156105
   /note="assembly-fragment"
misc_feature
156206..162048
   /note="assembly-fragment"
misc_feature
162149..165118
   /note="assembly-fragment"
misc_feature
165219..167480
   /note="assembly-fragment"
misc_feature
167581..169986
   /note="assembly-fragment"
misc_feature
170087..171909
   /note="assembly-fragment"
misc_feature
172010..172317
   /note="assembly-fragment"
misc_feature
172418..174663
   /note="assembly-fragment"
misc_feature
174764..176271
   /note="assembly-fragment"
misc_feature
176372..178022
   /note="assembly-fragment"
BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others
ORIGIN

```

Query Match 45.4%; Score 465.8; DB 2; Length 178022;
 Best Local Similarity 99.2%; Pred. No. 8e-97;
 Matches 489; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

QY 126 AGGTCCACCGGGGCCCGAGGGGTCTCAAGAGTGCACAGAGATCCACAGGACCCCTGGGCC 185
      |||||||
Db 123478 AGGTCCACCGGGGCCCGAGGGGTCTCAAGAGTGCACAGAGATCCACAGGACCCCTGGGCC 123537
QY 186 AACTGGCAACAGGACAGAAAGAGAGAAAGGGAGGAGCTGTGACACACTGGCCCTGGGG 245
      |||||||
Db 123538 AACTGGCAACAGGACAGAAAGAGAGAAAGGGAGGAGCTGTGACACACTGGCCCTGGGG 123597
QY 246 TGAGAGAGGCCCAATTGGACACAGCTGTGTCCTCCCGAGAGCTGTGGGCGCAAAAGATCTAA 305
      |||||||
Db 123598 TGAGAGAGGCCCAATTGGACACAGCTGTGTCCTCCCGAGAGCTGTGGGCGCAAAAGATCTAA 123657
QY 306 AGGCTCCAGGGGCCCGCAAGAGCTCCCGTGTTCCTCCCGAGAGCTCCCGTCAAGGGCC 365
      |||||||
Db 123658 AGGCTCCAGGGGCCCGCAAGAGCTCCCGTGTTCCTCCCGAGAGCTCCCGTCAAGGGCTC 123717
QY 366 CAGTGGGAGCCAGGCGCCCGGGGCCACACAGGCAAAAGAGAGCTCCCGGCCCTCAGGG 425
      |||||||
Db 123718 CAGTGGGAGCCAGGCGCCCGGGGCCACACAGGCAAAAGAGAGCTCCCGGCCCTCAGGG 123777
QY 426 CCCTCCTGGCTTCAGAGGACTTCAGGGACCGCTTGGGGAGCTGGGGTCTGTGACCTCG 485
      |||||||
Db 123778 CCCTCCTGGCTTCAGAGGACTTCAGGGACCGCTTGGGGAGCTGGGGTCTGTGACCTCG 123837
QY 486 GGGACTGGCAGGCTTGCTGGGGGTA-CCAGGATGTCACAGGCCCGCAAGGGGCCCGCGGCC 544
      |||||||
Db 123838 GGGACTGGCAGGCTTGCTGGGGGTACCCAGGATGTCACAGGCCCGCAAGGGGCCCGCGGCC 123897
QY 545 CTCCTGGCCATCAGAGGCGGTGTGTCCTCGGCCCTCGAGATGAGCCACCGCGGCAC 604
      |||||||
Db 123898 CTCCTGGCCATCAGAGGCGGTGTGTCCTCGGCCCTCGAGATGAGCCACCGCGGCAC 123956
QY 605 CGAGGACCAATGG 617
      |||||||
Db 123957 CGAGGACCAATGG 123969

```

RESULT 15

AP000939/c

LOCUS Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING

DEFINITION

DRAFT SEQUENCE, 32 unordered pieces.

ACCESSION

AP000939.3 GI:8119080

VERSION

HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

Homo sapiens DNA, clone:RP11-839023.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 188255)

AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

TITLE

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

JOURNAL

Published Only in Database (1999)

REFERENCE

2 (bases 1 to 188255)

AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

TITLE

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

JOURNAL

Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical

REFERENCE

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

AUTHORS

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

TITLE

Japan (E-mail:hattori@gsc.riken.go.jp,

JOURNAL

URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,

REFERENCE

On May 31, 2000 this sequence version replaced gi:6997751.

COMMENT

----- Genome Center

AUTHORS

Center: RIKEN Genomic Sciences Center(GSC)

TITLE

Center code: RIKEN

JOURNAL

Web site: http://hgp.gsc.riken.go.jp/

REFERENCE

Contact: hattori@gsc.riken.go.jp

AUTHORS

----- Project Information

TITLE

Center project name: HumDrafltl8

```
Center clone name: RP11-839023
----- Summary Statistics
Sequencing vector: PCR products: 100% of reads
Chemistry: Dye-terminator EF-merisham: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 171068 bases at least Q40
Consensus quality: 178865 bases at least Q30
Consensus quality: 182445 bases at least Q20
Insert size: 185155; sum-of-ctrls
Quality coverage: 4.41x in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1      16096 contig of 16096 bp in length
16197      33789 contig of 17593 bp in length
33890      51010 contig of 17121 bp in length
51111      63558 contig of 12448 bp in length
63659      75056 contig of 11398 bp in length
75157      84737 contig of 9581 bp in length
84838      90909 contig of 6072 bp in length
91010      105978 contig of 7606 bp in length
106079     113049 contig of 6971 bp in length
113150     119201 contig of 6052 bp in length
119302     126582 contig of 7281 bp in length
126683     132070 contig of 5388 bp in length
132171     137138 contig of 4968 bp in length
137239     143055 contig of 5817 bp in length
143156     148639 contig of 5484 bp in length
148740     153086 contig of 4347 bp in length
153087     153186 contig of 5735 bp in length
153187     158921 contig of 4153 bp in length
158922     159021 contig of 4153 bp in length
163175     163274 contig of 100 bp in length
163275     167066 contig of 3792 bp in length
167067     167166 contig of 100 bp in length
167167     170120 contig of 2954 bp in length
170121     170220 contig of 100 bp in length
170221     172430 contig of 2210 bp in length
172431     175598 contig of 3068 bp in length
175599     177118 contig of 1420 bp in length
177119     177218 contig of 1313 bp in length
177219     178531 contig of 100 bp in length
178532     178631 contig of 100 bp in length
178632     179857 contig of 1226 bp in length
179858     179957 contig of 100 bp in length
179959     181064 contig of 1107 bp in length
181065     181164 contig of 100 bp in length
181165     182735 contig of 1571 bp in length
182736     182835 contig of 100 bp in length
182836     183923 contig of 1088 bp in length
183924     184023 contig of 100 bp in length
184024     185810 contig of 1787 bp in length
185811     185910 contig of 100 bp in length
185911     187145 contig of 1235 bp in length
187146     187245 contig of 100 bp in length
187246     188255 contig of 1010 bp in length
Sequence updated (06-Jan-2000)
188255
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      16096 contig of 16096 bp in length
16097      16196 contig of 100 bp
16197      33789 contig of 17593 bp in length
33790      33889 contig of 100 bp
33890      51010 contig of 17121 bp in length
51011      51110 contig of 100 bp
51111      63558 contig of 12448 bp in length
63559      63658 contig of 100 bp
63659      75056 contig of 11398 bp in length
75057      75156 contig of 100 bp
75157      84737 contig of 9581 bp in length
84738      84837 contig of 100 bp
84838      90909 contig of 6072 bp in length
```

```
90910 91009: gap of 100 bp
91010 98615: contig of 7606 bp in length
98616 98715: gap of 100 bp
98716 105978: contig of 7263 bp in length
105979 106078: gap of 100 bp
106079 113049: contig of 6971 bp in length
113050 113149: gap of 100 bp
113150 119201: contig of 6052 bp in length
119202 119301: gap of 100 bp
119302 126582: contig of 7281 bp in length
126583 126682: gap of 100 bp
126683 132070: contig of 5388 bp in length
132071 132170: gap of 100 bp
132171 137138: contig of 4968 bp in length
137139 137238: gap of 100 bp
137239 143055: contig of 5817 bp in length
143056 143155: gap of 100 bp
143156 148639: contig of 5484 bp in length
148640 148739: gap of 100 bp
148740 153086: contig of 4347 bp in length
153087 153186: gap of 100 bp
153187 158921: contig of 5735 bp in length
158922 159021: gap of 100 bp
159022 163174: contig of 4153 bp in length
163175 163274: gap of 100 bp
163275 167066: contig of 3792 bp in length
167067 167166: gap of 100 bp
167167 170120: contig of 2954 bp in length
170121 170220: gap of 100 bp
170221 172430: contig of 2210 bp in length
172431 172530: gap of 100 bp
172531 175598: contig of 3068 bp in length
175599 175698: gap of 100 bp
175699 177118: contig of 1420 bp in length
177119 177218: gap of 100 bp
177219 178531: contig of 1313 bp in length
178532 178631: gap of 100 bp
178632 179857: contig of 1226 bp in length
179858 179957: gap of 100 bp
179959 181064: contig of 1107 bp in length
181065 181164: gap of 100 bp
181165 182735: contig of 1571 bp in length
182736 182835: gap of 100 bp
182836 183923: contig of 1088 bp in length
183924 184023: gap of 100 bp
184024 185810: contig of 1787 bp in length
185811 185910: gap of 100 bp
185911 187145: contig of 1235 bp in length
187146 187245: gap of 100 bp
187246 188255: contig of 1010 bp in length.
FEATURES
source
1..188255
Location/Qualifiers
1..188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1..16096
/feature="assembly_fragment"
16197..33789
/feature="assembly_fragment"
33890..51010
/feature="assembly_fragment"
51111..63558
/feature="assembly_fragment"
63659..75056
/feature="assembly_fragment"
75157..84737
/feature="assembly_fragment"
84838..90909
/feature="assembly_fragment"
91010..98615
/feature="assembly_fragment"
```

```
misc_feature      98716..105978
                  /note="assembly_fragment"
misc_feature      106079..113049
                  /note="assembly_fragment"
misc_feature      113150..119201
                  /note="assembly_fragment"
misc_feature      119302..126582
                  /note="assembly_fragment"
misc_feature      126683..132070
                  /note="assembly_fragment"
misc_feature      132171..137138
                  /note="assembly_fragment"
misc_feature      137239..143055
                  /note="assembly_fragment"
misc_feature      143156..148639
                  /note="assembly_fragment"
misc_feature      148740..153086
                  /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature      153187..158921
                  /note="assembly_fragment"
misc_feature      159022..163174

Query Match      42.6%; Score 437.4; DB 2; Length 188255;
Best Local Similarity 81.3%; Pred.No. 2.9e-90;
Matches 438; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 126 AGGTCCACCGGGGCCAGGGGTCCAGAGTGCACAGATCCAGGACCCCTGGGCC 185
      |||||||
Db 184461 AGGTCCACCGGGGCCAGGGGTCCAGAGTGCACAGATCCAGGACCCCTGGGCC 184402

QY 186 AACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCTCGGG 245
      |||||||
Db 184401 AACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCTCGGG 184342

QY 246 TGAGAGAGGCCCAATTGGACACAGCTGTCCCGGAGAGCTGGCGGCAAGATCTAA 305
      |||||||
Db 184341 TGAGAGAGGCCCAATTGGACACAGCTGTCCCGGAGAGCTGGCGGCAAGATCTAA 184282

QY 306 AGGCTCCAGGGGCCCAAGGCTCCCGTGTCCCTGGGAAAGCCCGCCCTCAGGGCCC 365
      |||||||
Db 184281 AGGCTCCAGGGGCCCAAGGCTCCCGTGTCCCTGGGAAAGCCCGCCCTCAGGGCCC 184222

QY 366 CAGTGGGACCCAGGCCCCCGGCCACACAGGCAAAAGAGGACTCCCGGCCCTCAGGG 425
      |||||||
Db 184221 CAGTGGGACCCAGGCCCCCGGCCACACAGGCAAAAGAGGACTCCCGGCCCTCAGGG 184162

QY 426 CCCTCCTGGCTTCCAGGACTTCAGGGCACCGTTGGGAGACCTGGGGTGCCTGGACTCG 485
      |||||||
Db 184161 CCCTCCTGGCTTCCAGGACTTCAGGGCACCGTTGGGAGACCTGGGGTGCCTGGACTCG 184102

QY 486 GGGACTGCCAGGCTTCTGGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGGCCC 545
      |||||||
Db 184101 GGGACTGCCAGGCTTCTGGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGGCCC 184042

QY 546 TCCTGCCCATCAGAGCGCGTGCCTTGGCCCTGCAGAAATGAGCAACCCCGGCACC 605
      |||||||
Db 184041 TCCTGCCCATCAGAGCGCGTGCCTTGGCCCTGCAGAAATGAGCAACCCCGGCACC 183982

QY 606 GGAAGACAATGGCTGCCGCCCTCAGTGAAGAACTCACAAGACAATGCTACTATTTT 664
      |||||||
Db 183981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNT 183923
```

Search completed: March 21, 2003, 06:43:48
Job time : 3104.5 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 02:00:05 ; Search time 36.1652 Seconds
(without alignments)
8700.356 Million cell updates/sec

Title: US-09-763-712a-1_COPY_670_1695
Perfect score: 1026
Sequence: 1 atgcacacagattgatgag.....cagctactcctcgtcattca 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.6	11.9	810	1	US-08-642-255-60 Sequence 60, Appl
2	118	11.5	756	1	US-08-642-255-50 Sequence 50, Appl
3	113.6	11.1	3181	1	US-08-655-086-1 Sequence 1, Appl
4	106	10.3	432	1	US-08-642-255-48 Sequence 48, Appl
5	105	10.3	1608	4	US-09-029-348-19 Sequence 19, Appl
6	100	9.7	5102	1	US-08-484-168-1 Sequence 1, Appl
7	99.6	9.7	1416	1	US-07-621-0916-1 Sequence 1, Appl
8	99.6	9.7	1416	2	US-08-399-889-1 Sequence 1, Appl
9	99.6	9.7	1416	3	US-09-167-364-1 Sequence 1, Appl
10	99.6	9.7	1416	4	US-09-439-897-1 Sequence 1, Appl
11	99.4	9.7	1868	1	US-08-392-3678-1 Sequence 1, Appl
12	99.4	9.7	1868	3	US-08-893-467A-1 Sequence 1, Appl
13	99.4	9.7	2543	1	US-08-535-669-11 Sequence 11, Appl
14	99.4	9.7	2543	3	US-09-073-663-11 Sequence 11, Appl
15	97.6	9.5	4359	4	US-09-484-9708-4 Sequence 4, Appl
16	96.2	9.4	1572	4	US-09-297-269-39 Sequence 39, Appl
17	95.2	9.3	1560	2	US-08-794-795-5 Sequence 5, Appl
18	95.2	9.3	1560	4	US-08-249-200-5 Sequence 5, Appl
19	95.2	9.3	1703	4	US-08-794-795-1 Sequence 1, Appl
20	95.2	9.3	1703	2	US-09-249-200-1 Sequence 1, Appl
21	89	8.7	1881	4	US-09-029-348-20 Sequence 20, Appl
22	84.4	8.2	506	1	US-08-253-155A-61 Sequence 61, Appl
23	77	7.5	1560	4	US-09-453-702B-264 Sequence 264, App
24	77	7.5	61663	4	US-09-453-702B-62 Sequence 62, Appl
25	76.8	7.5	5552	3	US-08-155-888-1 Sequence 1, Appl
26	74.6	7.3	38584	4	US-09-453-702B-50 Sequence 50, Appl
27	73	7.1	503	4	US-09-297-269-40 Sequence 40, Appl

28	72.2	7.1	4031	1	US-08-159-784-1	Sequence 1, Appl
29	72.4	7.0	9827	4	US-09-453-702B-66	Sequence 66, Appl
30	71.4	7.0	392	4	US-09-404-879A-372	Sequence 372, Appl
31	70	6.8	513	4	US-09-134-001C-647	Sequence 647, Appl
32	70	6.8	585	4	US-09-134-001C-726	Sequence 726, Appl
33	70	6.8	810	4	US-09-134-001C-624	Sequence 624, Appl
34	70	6.8	3552	4	US-09-134-001C-693	Sequence 693, Appl
35	69	6.7	1347	4	US-09-140-804-1	Sequence 1, Appl
36	68.6	6.7	45175	4	US-09-453-702B-116	Sequence 116, Appl
37	67.8	6.6	3404	4	US-09-453-702B-94	Sequence 94, Appl
38	67.8	6.6	48908	4	US-09-453-702B-137	Sequence 137, Appl
39	67.2	6.5	729	4	US-09-140-804-10	Sequence 10, Appl
40	67	6.5	226	1	US-08-175-155-63	Sequence 63, Appl
41	67	6.5	226	1	US-08-477-509B-98	Sequence 98, Appl
42	67	6.5	226	1	US-08-642-255-96	Sequence 96, Appl
43	67	6.5	226	1	US-08-397-633A-16	Sequence 16, Appl
44	67	6.5	226	2	US-08-707-237A-70	Sequence 70, Appl
45	67	6.5	226	3	US-08-482-085B-98	Sequence 98, Appl

ALIGNMENTS

```
RESULT 1
US-08-642-255-60
: Sequence 60, Application US/08642255
: Patent No. 5773249
:
: GENERAL INFORMATION:
: APPLICANT: CAPPELLO, Joseph
: APPLICANT: FERRARI, Franco A.
: TITLE OF INVENTION: High Molecular Weight Collagen-Like
: TITLE OF INVENTION: Protein Polymers
: NUMBER OF SEQUENCES: 135
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/642,255
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: ROWLAND, Bertram I.
: REGISTRATION NUMBER: 20,015
: REFERENCE/DOCKET NUMBER: A55556-3/BIR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEX: 910 277299 FRT UR
: INFORMATION FOR SEQ ID NO: 60:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 810 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "synthetic"
:
: US-08-642-255-60
:
: Query Match 11.9% Score 121.6; DB 1; Length 810;
: Best Local Similarity 53.9%; Pred. No. 1.9e-23;
: Matches 250; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
:
: QY 120 ACTCAAGGTCCACCGGGCCCGAGGGGTCCAGAGGTGCAGAGGATCCAGGACCC 179
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db	102	ACCAACGGGTGCACCTGAGCCAGGCGGGTCCGCTGTGATCTTAGAGTGAACCAAGACCGCC	161
QY	180	TGCGCCCAACTGGCGCAACAAGGACAAAAGAGAGAAGGGGAGGCTGTGACACTGTGGCC	239
Db	162	TGTGTCTCCGGGACCTGTCAGGCGCACCGGGGTACCGTGTGGGATCCGGGACCAACGGGTGC	221
QY	240	TGCGGTGATAGAGAGGCGCCAAATTGAGCAGCTGTCCCTCCGGAGAGGTGAGCGCAAG	289
Db	222	ACCTGCGCCACACGGGTTCGGCTGTGATCTAAGAGGTGACCAAGACCGCTGTGTCCGGG	281
QY	300	ATCTAAAGGCTCCAGAGGCGCCCAAAAGCTCCCGTGTTCCTCTGGAGACCGCGCCTCA	359
Db	282	ACTGTCAAGGCCACACGGGTAGCCGTGGCGATCCGAGACCAACGGGTTACACTGTGCGCAGC	341
QY	380	GGGCGCCGATGGGGAGCCACGAGCGCCCGGGGCGCCACACAGGCAANAAGGACTCCCGGGCC	419
Db	342	GGGTCCGCTGGATCTTAGAGTGTGACCAAGACGCCCTGTGTGTCTCCGGGACTGTGAAGGCC	401
QY	420	TCAGGGCCCTCTGTGCTTCACAGGACTTCACAGGACACCGTTGGGAGAGCTGTGGGTCCTGG	479
Db	402	ACCGGATGACCGGTGGCGATCCGGGAGCCACACGCGGTGCACATCTGGGCCCAAGGGTCCGCTTG	461
QY	480	ACCTTGGGGACTGCCAGGCTTGCTGGGGGTACCAAGGATGCGCAGGCCCCAAGGGCCCCC	539
Db	462	ATCTAGAGTGAACCCAGGACCGCTGTGTGTCTCCGGGACCTTCAGAGGCCACACGGGTAGCCG	521
QY	540	CGGGCCCTCTGGCCCATCAGGAGGGGTGTGTCCTCTGGCCCTGC	583
Db	522	TGGCGATCCGGGACCAACCGGTTGCACTTGACCTACGCGGCTCCG	565

RESULT 2
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249

GENERAL INFORMATION:

APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco

;	TITLE OF INVENTION:	High Molecular Weight Collagen-Like
;	TITLE OF INVENTION:	Protein Polymers

```

; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400

;
CITY: San Francisco
;
STATE: California

COUNTRY: USA
ZIP: 94111-4187

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
```

```

;
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
;

```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
:

```

APPLICATION NUMBER: US/08/642,255
FILING DATE:

CLASSIFICATI
ATTORNEY/AGENT

NAME: ROWLAND, Bertram I
REGISTRATION NUMBER: 30

REGISTRATION NUMBER: 20/013
REFERENCE/DOCKET NUMBER: A55556-3/BIR
FEE/COMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8111
TELEX: 910 277299 FHT UR

```

; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 756 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
;

```

US-08-642-255-50

Query Match	11.5%	Score 118;	DB 1;	Length 756;
Best Local Similarity	54.3%;	Pred. No. 1.7e-22;		
Matches 238;	Conservative	0;	Mismatches 200;	Indels 0;
				Gaps 0;

QY	127	GGTCCACCGGGCCCCCAGGAGGTGCACAAAGAGTGTACACAGAGATGCCAGGAGGACCCCTGGGCCA	186
Db	289	GGTCCACCGGGTGTCTCCGGGACCTGCAGGCCCGCCGACAGTGTGCTGTGACCCGGCTGTGTCCA	348
QY	187	ACTGGCAACAAGGACGACAGAAAGACAGAGAGAGGGGAGGACCTGTGACCACTGTGCCCCGCGGT	246
Db	349	CCGGGTGTCTCCGGGACCTGCAGCCGCCCCACAGGTGTGCGCTGTGACCGGTGTGTGTCCACCGCGGT	408
QY	247	GAGAGAGGGCCCAATTGGACAGCTGTGTGTCCCGGAGACAGGTGTGGCGSCAAAGGATCTAA	306
Db	409	GCTCCGGGACCTGCAGAGCCCGCCACAGGTGTGCGCTGTGACCGGCTGTGTCCACCGGGTGTCTCCG	468
QY	307	GGCTCCACGAGGCCCCCAAAAGGCTCCCGTGTGTTCCTCTGGAAAGCCCGGCTCTCAGGGGCCCC	366
Db	469	GGACCTCGAGGCCCGCCAGGAGTGTGCTGTGACCCGGCTGTGTCCACCGGGTGTCTCCGGGACCT	528
QY	367	AGTGGGGAACCGACGCCCCCGCGGCCACACAGGCAAAAGGAGTCTCCCGGCTCTCAGGCG	426
Db	529	GCAAGCCCGCGAGGTGTGGCCCTGGACCGGCTGTGTCAACCGGGTGTCTCCGGGACCTGCAGCG	588
QY	427	CCCTCTGTGCTTCAGAGGACTTCAGAGGACACGGTGTGGGAGGCTGTGGGAGTGTGTGACCTCGG	486
Db	589	CCGGCAGTGTGGCTTGTGACCGCGCTGTGTTCACCGGGTGTCTCCGGGACTGTGCAAGGCCCGCCA	648
QY	487	GGACTGTCCAGGCTTGTGGGGTACCAAGGCATGTCCAGGCCCCCAAGGGGCCCGCCCGGCTCT	546
Db	649	GGTGTGCGCTGTGACCGGCTGTGTGTCCACCGGGTGTCTCCGGGACTGTGCAAGGCCCGCCAGGTGCC	708
QY	547	CGTGGCCCATCAGGAGCGG	564
Db	709	CATGGCCCATCAGGAGCGG	726

RESULT 3
US-08-65

; Sequence 1, Application US/086550866

; Patent NO. 5821089
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOTT A
; ;
APPLICANT: BUECHTER, DOUGLAS
; ;

;
; APPLICANT: ZHANG, GUANGHUI
;
; APPLICANT: CONNOLLY, KEVIN

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD

CITY: UNIONDALE
STATE: NY

COUNTRY: US
ZIP: 11553

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

; COMMENT AFFILIATION DATA:
; APPLICATION NUMBER: US/08/655,086
; FILING DATE: 03 JUN 2005

```

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S.

```

; REGISTRATION NUMBER: 32,063
; REFERENCE/DOCKET NUMBER: 203-1632

```

TELECOMMUNICATION INFORMATION
TELEPHONE: 516-228-8484

TELEFAX: 516-228-8516
; INFORMATION FOR SEQ ID NO: 1:

Query Match 10.3%; Score 106; DB 4; Length 1608;
Best Local Similarity 52.8%; Pred. No. 4.1e-19;
Matches 229; Conservative 0; Mismatches 205; Indels 0; Gaps 0;


```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 972-1400
: TELEFAX: (212) 370-1622
: TELEX: 236268
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1416 base pairs
: TYPE: Nucleic acid
: STRANDEDNESS: Double
: TOPOLOGY: Linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Calf
: STRAIN: Unknown
: INDIVIDUAL ISOLATE: Unknown
: DEVELOPMENTAL STAGE: Unknown
: CELL TYPE: Whole kidney
: IMMEDIATE SOURCE:
: LIBRARY: Bovine lens cDNA
: CLONE: KMC15
: POSITION IN GENOME: No. 5424408 known
: CHROMOSOME/SEGMENT: No. 5424408 known
: PUBLICATION INFORMATION: No. 5424408e
: US-07-621-091G-1

```

```

Query Match          9.7%; Score 99.6; DB 1; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

```

QY	127	GGTCCACCGGGCCCCCAGGGGTCCAGAGGTGCAGAGGATCCACAGGGACCCCGCTGGCCCA	186
Db	283	GGAGCACACAGGTCTCCCGGTACGCCAGGGGCCAGAGGTGATCCTGGATTCTATGGATT	342
QY	187	ACTGGCAACAAGGCACAGAAGCAGAGAAGGGGAGCCCTGGACCCTCGCCCTCGGGT	246
Db	343	CCAGGCATGAAGGAGAGGTAATTTCAGGATTTCCAGGACCCTCGACTCCAGG	402
QY	247	GAGAGAGGCCAATTCGACACGTGTCTCCCGGAGAGCGTGGGGCAAAAGATCTAAA	306
Db	403	CAAAGTGGACCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC	462
QY	307	GGCTTCCCAGGGCCCCAAAGGCTCCCGTGTCTCCCTGGGAAGCCGGCCCTCAGGGCCCC	366
Db	463	TCCCTTCAGGAAGCCAGGCCACCTGGTTTCAGCTGGAGAACCGGAGTCAAGAGGAA	522
QY	367	AGTGGGACCCAGGCCCCCCGGCCCCACAGGCAAAAGAGGAGTCTCCCGGCCCTCAGGGC	426
Db	523	CCCGGGCCCCAGGACACCCAGAGATCCAGGACCTGTGGGCCAAAAGTAAACCAGG	582
QY	427	CTCTCTGGCTTCAGGACTTCAGGCACCGTTTGGGAGCGCTGGGGTGCCTGGACCTCGG	486
Db	583	GAGGATGGTCCACAGGAATCTCTGGACCAACTTGGAGAAAAGGCACAAAGGTTGTAAA	642
QY	487	GGACTGCCAGGCTTGCTCGGGGTACCAAGGCATGCCAGGGCCCAAGGGCCCCCGGCCCT	546
Db	643	GGAGAGCAAGGACCACTGGATCCGATGGCCTGCCAGGCTTGAGGGGAAACCTGGAGAC	702
QY	547	CTTGGCCCATCAGAGCGGTGG	568
Db	703	ACTGGACCACCTGCACAGGGG	724

```

RESULT 8
US-08-399-889-1
; Sequence 1, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain T

```

```

: FILE REFERENCE: 951263A
: CURRENT APPLICATION NUMBER: US/08/399, 889B
: CURRENT FILING DATE: 1995-03-07
: EARLIER APPLICATION NUMBER: 07/621091
: EARLIER FILING DATE: 1990-11-30
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1

```

```

; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-08-399-889-1

```

Query Match 9.7%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. NO. 2e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy	127	GGTCCACCGGGCCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCACAGGGACCCCTCGGCCCA	186
Db	283	GGAGCACAGGTCTCCCGGTGACGACAGGGGCCAGAGGTGATCCTGGATTCTATGGATT	342
Qy	187	ACTGGCAACAAGGACAGAAAGCAGAGAGAGGGGAGCCTGGACCACTGCGCCTCGCGGT	246
Db	343	CCAGGCATGAAGGGAAGAAAGGGTAATTCAAGATTTCAGAGCACCACTGACCTCCAGGG	402
Qy	247	GAGAGAGGCCCAATTGGACACAGCTGTGCCCCCGAGAGCGTGGCGGCAAAAGSATCTAA	306
Db	403	CAAAGTGGACCAAAAGSACCACCTGGGTACGTGGAGAGCCTGGCACAGTGAAGATCATC	462
Qy	307	GGCTCCAGGGCCCCCAAGAGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAGGGCCCC	366
Db	463	TCCTCTCCAGGAAGCCAGGCCACCTGGTTCAGCTGGAGAACCAGGATGCAAGAGAA	522
Qy	367	AGTGGGACCCAGGCCCCCGGGCCCCACAGGCAAAAGAGGAGCTCCCGGCCCTCAGGGC	426
Db	523	CCGGGGCCCCAGGACCACAGGAGATCCAGGACCCTGTGGGCCAAAAGSTAAACCCAGG	582
Qy	427	CTCTCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGAGCGCTGGGTGCTGGACCTCGG	486
Db	583	GAGGATGGTCCACAGGAACCTCTGACCAACTCTGGACCAAGTGGAGAAAAGGGCAACAAAGGTGTGAA	642
Qy	487	GGATGCGACGGCTGTCTGGGTACCAAGGCATGCCAGGCCCCCAAGGGCCCCCGCGCCCT	546
Db	643	GGAGAGCAAGSACCACTGATCCGATGGCCTGCCAGGCTTGAAGGGGAACCTGGAGAC	702
Qy	547	CCTGGGCCCATCAGAGCGGTGG	568
Db	703	ACTGGACCACTGCAGCGGG	724

```

RESULT 9
US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263b
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf

```

FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1416)
US-09-167-364-1

Query Match 9.7%; Score 99.6; DB 3; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 127 GTTCCACCGGGCCCGAGAGGTTCACAGAGGTGACAGAGGATCCAGGACCCCTCGGCCCA 186
DB 283 GGAGCACCAAGTGTCCCGGTGACGACGGGCGCAGAGGTGATCTGTGATTTATGATTT 342
QY 187 ACTGCCAACAAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACACCTGGCCCTCGGGGT 246
DB 343 CCAGGCATCAAAAGGAAGAGGTAAATTCAGGATTTCCAGGACCACTGTGACCTCAGG 402
QY 247 GAGAGAGGCCCAATTTGGACCACTGTGTCCTCCCGGAGAGCGTGGGGGCAAGAGTCTAAA 306
DB 403 CAAAGTGGACCAAAAGGACCACTGTGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC 462
QY 307 GCTCCACGGGCCCCAAAGGCTCCGTTGTTCCCTGGGAAGCCCGCCCTCAGGCCCC 366
DB 463 TCCTTCCAGGAAGCCCGAGGCCACCTGTGTTCACTGAGGAGACCAAGGATGCAAGAGAA 522
QY 367 AGTGGGACCCAGGCCCCCGGCCCCCACCAGGCAAAAGAGGACTTCCCGGCCCTCAGGGC 426
DB 523 CCGGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCAAAAGGTAACACAGG 582
QY 427 CCTCTGGGCTTCAGGGACCTTCAGGGACCGTTGGGGAGCCTGGGGTGCCTGACCTCGG 486
DB 583 GAGGATGTTCACAGGAACCTCTTGACCACTGGAGAAAAAGGCACAAAGGTTGTATAA 642
QY 487 GGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCT 546
DB 643 GGAGAGCAAGGACCACTGGATCCGATGCGCTGCCAGGCTTGAAGGGGAAACCTTGGAGAC 702
QY 547 CTGTGCCCATCAGGACGGTGG 568
DB 703 ACTGGACCACTGCAGCAGGG 724

RESULT 10
US-09-439-897-1
Sequence 1, Application US/09439897
Patent No. 6277558
GENERAL INFORMATION:
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1416)
US-09-439-897-1

Query Match 9.7%; Score 99.6; DB 4; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 127 GTTCCACCGGGCCCGAGAGGTTCACAGAGGTGACAGAGGATCCAGGACCCCTCGGCCCA 186
DB 283 GGAGCACCAAGTGTCCCGGTGACGACGGGCGCAGAGGTGATCTGTGATTTATGATTT 342
QY 187 ACTGCCAACAAAGGACAGAAAGGAGAGGGGAGGCTGGACCACTGTGCGCCCTCGGGT 246

DB 343 CCAGGCATGAAAGGGAAGAGGTAAATTCAGGATTTCCAGGACCACTTGACCTCCAGGG 402
QY 247 GAGAGAGGCCCAATTTGGACCACTGGTCCCGCCCGGAGAGCGTGGCGGCAAAAGGATCTAAA 306
DB 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGCACACATGAAGATCATC 462
QY 307 GCTCCACGGGCCCCAAAGGCTCCGTTGTTCCCTGGGAAGCCCGGCCCTCAGGCCCC 366
DB 463 TCCTTCCAGGAAGCCCGAGGCCACCTGTGTTCACTGAGTGGAGAACCAAGGATGCAAGAGAA 522
QY 367 AGTGGGACCCAGGCCCCCGGCCCCCACCAGGCAAAAGAGGACTTCCCGGCCCTCAGGGC 426
DB 523 CCGGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCAAAAGGTAACACAGG 582
QY 427 CCTCTGGGCTTCAGGGACCTTCAGGGACCGTTGGGGAGCCTGGGGTGCCTGACCTCGG 486
DB 583 GAGGATGTTCACAGGAACCTCTTGACCACTGGAGAAAAAGGCACAAAGGTTGTATAA 642
QY 487 GGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCT 546
DB 643 GGAGAGCAAGGACCACTGGATCCGATGCGCTGCCAGGCTTGAAGGGGAAACCTTGGAGAC 702
QY 547 CTGTGCCCATCAGGACGGTGG 568
DB 703 ACTGGACCACTGCAGCAGGG 724
RESULT 11
US-08-392-367B-1
Sequence 1, Application US/08392367B
Patent No. 5691197
GENERAL INFORMATION:
APPLICANT: Tryggvason, Karl
APPLICANT: Elomaa, Outi
APPLICANT: Kangas, Maarit
TITLE OF INVENTION: An Insolated DNA Sequence For a
Patent No. 5691197
TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
TITLE OF INVENTION: a Collagenous Domain and the
TITLE OF INVENTION: Polypeptide Chain Encoded by
TITLE OF INVENTION: such a Sequence
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fay, Sharpe, Beall, Fagan,
ADDRESSEE: Minnich & McKee
STREET: 1100 Superior Avenue
STREET: Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: U.S.A.
ZIP: 44114-2518
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch,
MEDIUM TYPE: 720 Kb storable
COMPUTER: IBM PS/2, Model 35 SX
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,367B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard J.
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 2 009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1868 base pairs
TYPE: Nucleic acid

```
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 5691197 relevant
; ANTI-SENSE: No. 5691197 relevant
US-08-392-367B-1

Query Match          9.7%; Score 99.4; DB 1; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.6e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGGTCCACCGGGCCCCAGAGGTCCAAAGGTGACAGAGATCCACAGGAGCCCTCGC 183
Db 721 CCAGGTGTCCAAGGACCCAGGCGCCACCAGGCAAGGAGGAGGCGCTCCAGGGA 780
Qy 184 CCAACTGGCAACAAGGACAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 243
Db 781 CTTAGGGGTGCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy 244 GGTGAGAGAGGCCCCAAATTTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 303
Db 841 GGCAGCAAGAGTGACATAGGTCTCACTGGCCCCCAAGGGGGAACATGGCCACCAAGGGAGAC 900
Qy 304 AAAGGCTCCAGGCGCCCCAAAGGCTCCCGTGTTCCTTGGGAAGCCCGGCTCAGGCG 363
Db 901 AAAGGGACCTAGGCGCTTCCAGGAAACAAGGGGAGACATGGGCATGAAGGGAGACACGCGG 960
Qy 364 CCCAGTGGGAGCCAGGCGCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
Db 961 CCCATGGGTCCTCTGGAGCTCAGGAGGAGTAAGGTGATGCTGGAACACGAGGCTACCA 1020
Qy 424 GGCCCTCTCGCTTCCAGGAGCTTCAGGCGACCCCTTTGGGAGCCCTGGGGTGCCTGACCT 483
Db 1021 GGTTCGTGGTATCTCCAGGAGTCAAAAGTGACCAAGGAAACCTGGAGTGCAGGTTGT 1080
Qy 484 CGGGAGCTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCCCAAGGGCCCCCGCGC 543
Db 1081 CCAGGCCCTCAAGGTGCACAGGACTTTCAGGTGCCAAGGTTGACCCAGGAGCAGCACTGTT 1140
Qy 544 CTCTCTGCCATCAGGAGCGGTGG 568
Db 1141 CTCTCTGCCAGCAGGACCCCCCG 1165

RESULT 12
US-08-893-467A-1
; Sequence 1, Application US/08893467A
; Patent No. 6063901
; GENERAL INFORMATION:
; APPLICANT: Tryggevason, Karl
; APPLICANT: Elomaa, Outi
; APPLICANT: Kangas, Maarit
; TITLE OF INVENTION: An Isolated DNA Sequence For a
; Patent No. 6063901
; TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
; TITLE OF INVENTION: a Collagenous Domain and the
; TITLE OF INVENTION: Polypeptide Chain Encoded by
; TITLE OF INVENTION: such a Sequence
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 720 kb storable
; COMPUTER: IBM PS/2, Model 35 SX
; OPERATING SYSTEM: DOS 5.0
```

```
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,467A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 2 009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 6063901 relevant
; ANTI-SENSE: No. 6063901 relevant
US-08-893-467A-1

Query Match          9.7%; Score 99.4; DB 3; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.6e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGGTCCACCGGGCCCCAGAGGTCCAAAGGTGACAGAGATCCACAGGAGCCCTCGC 183
Db 721 CCAGGTGTCCAAGGACCCAGGCGCCACCAGGCAAGGAGGAGGAGGAGGAGGAG 780
Qy 184 CCAACTGGCAACAAGGACAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAG 243
Db 781 CTTAGGGGTGCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy 244 GGTGAGAGAGGCCCCAAATTTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 303
Db 841 GGCAGCAAGAGTGACATAGGTCTCACTGGCCCCCAAGGGGGAACATGGCCACCAAGGGAGAC 900
Qy 304 AAAGGCTCCAGGCGCCCCAAAGGCTCCCGTGTTCCTTGGGAAGCCCGGCTCAGGCG 363
Db 901 AAAGGGACCTAGGCGCTTCCAGGAAACAAGGGGAGACATGGGCATGAAGGGAGACACGCGG 960
Qy 364 CCCAGTGGGAGCCAGGCGCCCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
Db 961 CCCATGGGTCCTCTGGAGCTCAGGAGGAGTAAGGTGATGCTGGAACACGAGGCTACCA 1020
Qy 424 GGCCCTCTCGCTTCCAGGAGCTTCAGGCGACCCCTTTGGGAGCCCTGGGGTGCCTGACCT 483
Db 1021 GGTTCGTGGTATCTCCAGGAGTCAAAAGTGACCAAGGAAACCTGGAGTGCAGGTTGT 1080
Qy 484 CGGGAGCTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCCCAAGGGCCCCCGCGC 543
Db 1081 CCAGGCCCTCAAGGTGCACAGGACTTTCAGGTGCCAAGGTTGACCCAGGAGCAGCACTGTT 1140
Qy 544 CTCTCTGCCATCAGGAGCGGTGG 568
Db 1141 CTCTCTGCCAGCAGGACCCCCCG 1165

RESULT 13
US-08-555-669-11
; Sequence 11, Application US/08555669
; Patent No. 5773248
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
```

;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/555.669
;; FILING DATE: 13-NOV-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Halluin, Albert P.
;; REGISTRATION NUMBER: 25/227
;; REFERENCE/DOCKET NUMBER: 8389-030
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-854-3660
;; TELEFAX: 415-854-3694
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2543 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 47..2098
US-08-555-669-11

Query Match 9.7%; Score 99.4; DB 1; Length 2543;
Best Local Similarity 51.5%; Pred. No. 3e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 124 CAAGGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGTGACAGGGAGCCCTGGCCCTGGC 183
DB 584 CCAGTCCCGGAGGAGTCCGGAATGCCAGGGTTCAGAGGACCCACTGGCTACAAAGGC 643
QY 184 CCAACTGGCAACAAGGACACAAAGAGAGAGAGAGGGGAGCCCTGGACACCTGGCCCTGGG 243
DB 644 GAGCAGGGGGAAGTCCGGAAGGACGGCGAGAGAGGTGACCTGGCCCTGGCCCTGGCCCGCC 703
QY 244 GGTGAGAGAGGCCCAATTTGGACACGTGCTCCCGGAGAGCGTGGCGGCAAAAGATCT 303
DB 704 GCCTCCCGGGCAGCGTGGGGCTGAGGGCCCGCGGGGATTCAGAGGACTGCCAGGGCCA 763
QY 304 AAAGGCTCCAGGGCCCCAAAGGCTCCCGTCCCGGAGAGCGTGGCGGCAAAAGATCT 363
DB 764 CTCGGGCCCCCTGGGAGCCGGGGTCCCATTTGGTTCCGAGGACCCCTGGCCCTGGC 823
QY 364 CCCAGTGGGACCCAGGGCCCCCGGGCCACCAAGAGAGGGGACTCCCGGCCCCCTAG 423
DB 824 GCGCTCCGGAAGCGGGTGACCGAGGAGAGGGGCCCAAGAGGTTCGCGGCCCCAAG 883
QY 424 GCGCTCTCTGGCTTCAGGGACTTCAGGGACCGTTCGGGAGCCCTGGGGTGCCTGGACCT 483
DB 884 GTGACCTCGGACAGCTTGGTCCCAAGGAACCCCGGAGTGGCGGGCCAGAGGAGAG 943
QY 484 CGGGAGCTGCCAGGTTCCTTGGGTACCGCATGCCAGGCCCAAGGGCCCCCCCCCGC 543
DB 944 CGGGGATGCCAGGAAGGACGGCCAGAAATGCGTGCAGGACTCGATGGCCAGAAGGA 1003
QY 544 CTCCTGCCCCATCAGGACGGGTGG 568
DB 1004 GAGGCTGGTTCGCAACGGTGTCTCCGG 1028

RESULT 14
US-09-073-663-11
; Sequence 11, Application US/09073663

;; Patent No. 6127523
;; GENERAL INFORMATION:
;; APPLICANT: Brewton, Richard G.
;; APPLICANT: Mayne, Richard
;; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
;; NUMBER OF SEQUENCES: 32
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: McGregor & Adler, LLP
;; STREET: 8011 Candle Lane
;; CITY: Houston
;; STATE: Texas
;; COUNTRY: USA
;; ZIP: 77071
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: Apple Macintosh
;; OPERATING SYSTEM: Macintosh OS 8.1
;; SOFTWARE: Microsoft Word for Macintosh
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/073.663
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Benjamin Aaron Adler, Ph.D., J.D.
;; REGISTRATION NUMBER: 35,423
;; REFERENCE/DOCKET NUMBER: D5913D
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (713) 777-2321
;; TELEFAX: (713) 777-6908
;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2543 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 47..2098
US-09-073-663-11

Query Match 9.7%; Score 99.4; DB 3; Length 2543;
Best Local Similarity 51.5%; Pred. No. 3e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 124 CAAGGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGTGACAGGGAGCCCTGGC 183
DB 584 CCAGTCCCGGAGGAGTCCGGAATGCCAGGGTTCAGAGGACCCACTGGCTACAAAGGC 643
QY 184 CCAACTGGCAACAAGGACACAAAGAGAGAGAGGGGAGCCCTGGACACCTGGCCCTGGG 243
DB 644 GAGCAGGGGGAAGTCCGGAAGGACGGCGAGAGAGGTGACCTGGCCCTGGCCCTGGCCCGCC 703
QY 244 GGTGAGAGAGGCCCAATTTGGACACGTGCTCCCGGAGAGCGTGGCGGCAAAAGATCT 303
DB 704 GCCTCCCGGGCAGCGTGGGGCTGAGGGCCCGCGGGGATTCAGAGGACTGCCAGGGCCA 763
QY 304 AAAGGCTCCAGGGCCCCAAAGGCTCCCGTCCCGTCCCGGAGAGCGTGGCGGCAAAAGATCT 363
DB 764 CTCGGGCCCCCTGGGAGCCGGGGTCCCATTTGGTTCCGAGGACCCCTGGCCCTGGC 823
QY 364 CCCAGTGGGACCCAGGGCCCCCGGGCCACCAAGAGAGGGGACTCCCGGCCCCCTAG 423
DB 824 GCGCTCCGGAAGCGGGTGACCGAGGAGAGGGGCCCAAGAGGTTCGCGGCCCCAAG 883
QY 424 GCGCTCTCTGGCTTCAGGGACTTCAGGGACCGTTCGGGAGCCCTGGGGTGCCTGGACCT 483
DB 884 GTGACCTCGGACAGCTTGGTCCCAAGGAACCCCGGAGTGGCGGGCCAGAGGAGAG 943
QY 484 CGGGAGCTGCCAGGTTCCTTGGGTACCGCATGCCAGGCCCAAGGGCCCCCCCCCGC 543
DB 944 CGGGGATGCCAGGAAGGACGGCCAGAAATGCGTGCAGGACTCGATGGCCAGAAGGA 1003

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 156.339 Seconds
(without alignments)
5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_1695

Perfect score: 1026

Sequence: 1 atgcacaaagatttgatgag.....cagtactgtcatctgcatta 1026

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 38737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1024.4	99.8	2930	10	US-09-745-763-198
2	111	10.8	6728	10	US-09-954-456-782
3	111	10.8	6728	10	US-09-880-107-3946
4	110.6	10.8	1133	9	US-09-924-340-57
5	110.6	10.8	1133	9	US-09-922-600A-57
6	109.4	10.7	2691	10	US-09-925-302-64
7	107.2	10.4	2192	10	US-09-925-301-42
8	107	10.4	5086	10	US-09-880-107-3947
9	107	10.4	5145	9	US-09-925-299-206
10	107	10.4	5145	10	US-09-925-299-206
11	107	10.4	5432	12	US-10-044-090-22
12	106.8	10.4	4908	9	US-10-001-887-33
13	106	10.3	2542	9	US-09-954-531-961
14	106	10.3	2542	10	US-09-964-824A-255
15	105.4	10.3	5416	10	US-09-954-456-786
16	105.4	10.3	5416	10	US-09-880-107-2094
17	105.2	10.3	6158	10	US-09-919-497-6
18	105.2	10.3	6158	10	US-09-954-456-762
19	104.2	10.2	4149	12	US-10-044-090-104

20	102.4	10.0	1486	10	US-09-925-302-247	Sequence 247, App
21	101.6	9.9	3226	10	US-09-954-456-725	Sequence 725, App
22	101.4	9.9	3690	12	US-10-044-090-448	Sequence 448, App
23	98.6	9.6	2158	9	US-10-001-835-98	Sequence 98, Appl
24	97.8	9.5	2520	10	US-09-880-107-3685	Sequence 3685, Ap
25	95.8	9.3	1485	10	US-09-925-302-246	Sequence 246, App
26	95.2	9.3	1797	9	US-09-978-295A-613	Sequence 613, App
27	95.2	9.3	1797	9	US-09-978-697-613	Sequence 613, App
28	95.2	9.3	1797	9	US-09-978-192A-613	Sequence 613, App
29	95.2	9.3	1797	9	US-09-999-832A-613	Sequence 613, App
30	95.2	9.3	1797	9	US-09-978-189-613	Sequence 613, App
31	95.2	9.3	1797	9	US-10-174-590-331	Sequence 331, App
32	95.2	9.3	1797	9	US-10-176-758-331	Sequence 331, App
33	95.2	9.3	1797	9	US-10-175-737-331	Sequence 331, App
34	95.2	9.3	1797	9	US-10-173-706-331	Sequence 331, App
35	95.2	9.3	1797	9	US-10-175-738-331	Sequence 331, App
36	95.2	9.3	1797	9	US-10-175-752-331	Sequence 331, App
37	95.2	9.3	1797	9	US-10-176-482-331	Sequence 331, App
38	95.2	9.3	1797	9	US-10-176-757-331	Sequence 331, App
39	95.2	9.3	1797	9	US-10-176-913-331	Sequence 331, App
40	95.2	9.3	1797	9	US-10-180-552-331	Sequence 331, App
41	95.2	9.3	1797	9	US-10-180-557-331	Sequence 331, App
42	95.2	9.3	1797	9	US-10-173-700-331	Sequence 331, App
43	95.2	9.3	1797	9	US-10-174-572-331	Sequence 331, App
44	95.2	9.3	1797	9	US-10-174-579-331	Sequence 331, App
45	95.2	9.3	1797	9	US-10-174-582-331	Sequence 331, App

ALIGNMENTS

RESULT 1

US-09-745-763-198
; Sequence 198 Application US/09745763
; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Treacy, Maurice
Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS: ENCODING THEM

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 198:

SEQUENCE CHARACTERISTICS:

LENGTH: 2930 base pairs

TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match 99.8%; Score 1024.4; DB 10; Length 2930;
Best Local Similarity 99.9%; Pred. No. 1.7e-271;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAACAAAGTTTATGATAGGTCGAGGTAGACACTGAAGTAGCCCAACTTATCACTGATT 60
DB 1266 ATGCAACAAAGTTTATGATAGGTCGAGGTAGACACTGAAGTAGCCCAACTTATCACTGATT 1325
QY 61 ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGTGTCACTCATCAAGAAATTTTACAATA 120
DB 1326 ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGTGTCACTCATCAAGAAATTTTACAATA 1385
QY 121 CTACAAAGTTCACCGGGCCCCAGGGTCCAAGAGTGCAGAGGATCCAGAGGACCCCT 180
DB 1386 CTACAAAGTTCACCGGGCCCCAGGGTCCAAGAGTGCAGAGGATCCAGAGGACCCCT 1445
QY 181 GCCCAACTGGCAACAAGGACAGAAAGGAGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 240
DB 1446 GCCCAACTGGCAACAAGGACAGAAAGGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 1505
QY 241 GCGGGTAGAGAGGGCCCAATTTGACACAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGA 300
DB 1506 GCGGGTAGAGAGGGCCCAATTTGACACAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGA 1565
QY 301 TCTAAAGGTCACAGGGCCCCAAAGGCTCCCTGGTTCCTTGGGAAGCCCGGCCCTCAG 360
DB 1566 TCTAAAGGTCACAGGGCCCCAAAGGCTCCCTGGTTCCTTGGGAAGCCCGGCCCTCAG 1625
QY 361 GCGCCACAGTGGGACCCAGGCCCCCGGCCCAACAGGCAAGAGAGGACTCCCGGCCCT 420
DB 1626 GCGCCACAGTGGGACCCAGGCCCCCGGCCCAACAGGCAAGAGAGGACTCCCGGCCCT 1685
QY 421 CAGGGCCCTCTCGGCTTCAGGGACTTTCAGGACACCGTTGGGGAGCCTGGGGTGGCTGA 480
DB 1686 CAGGGCCCTCTCGGCTTCAGGGACTTTCAGGACACCGTTGGGGAGCCTGGGGTGGCTGA 1745
QY 481 CCTCGGGGACTCCAGGCTTGGGTACAGGATGCCAGGCCCCAGGCCCCAGGCCCCCCC 540
DB 1746 CCTCGGGGACTCCAGGCTTGGGTACAGGATGCCAGGCCCCAGGCCCCAGGCCCCCCC 1805
QY 541 GCGCCTCTCTGGCCATCAGAGCGGTGTGTCCTCGCCCTCAGAAATGAGCAACCCCG 600
DB 1806 GCGCCTCTCTGGCCATCAGAGCGGTGTGTCCTCGCCCTCAGAAATGAGCAACCCCG 1865
QY 601 GCACCGGAGGACAATGGCTGCCCCGCTCCTGGAAGACTTCACAGACAAATGCTACTAT 660
DB 1866 GCACCGGAGGACAATGGCTGCCCCGCTCCTGGAAGACTTCACAGACAAATGCTACTAT 1925
QY 661 TTTTCAGTTGAGAAAGAAATTTTGAAGATGCAAGCTTTTCTGTGAGACAAAGTCTCA 720
DB 1926 TTTTCAGTTGAGAAAGAAATTTTGAAGATGCAAGCTTTTCTGTGAGACAAAGTCTCA 1985
QY 721 CATCTGTTTTCATAAAGACTAGAGAGAACAGCAATGGATAAAAAACAGATGTAGGG 780
DB 1986 CATCTGTTTTCATAAAGACTAGAGAGAACAGCAATGGATAAAAAACAGATGTAGGG 2045
QY 781 AGAGAGGCCACTGTGATCGGCTCAGACTCAGAGCGGTGAAATGAATGGAAGTGGCTG 840
DB 2046 AGAGAGGCCACTGTGATCGGCTCAGACTCAGAGCGGTGAAATGAATGGAAGTGGCTG 2105
QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTTGACAGCCCGGATTAAGTGGGTGAT 900
DB 2106 GATGGACATCTCCAGACTACAAAATTTGGAAGCTTGACAGCCCGGATTAAGTGGGTGAT 2165
QY 901 GGCCATGGGCCAGGAGAGACTGTCTGGTTGATTTATGCTGGGAGTGGAAACGATTTC 960
DB 2166 GGCCATGGGCCAGGAGAGACTGTCTGGTTGATTTATGCTGGGAGTGGAAACGATTTC 2225

QY 961 CAATGTGAAGACGTCAATAACTTTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTATCT 1020
DB 2226 CAATGTGAAGACGTCAATAACTTTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTATCT 2285
QY 1021 GCATTA 1026
DB 2286 GCATTA 2291

RESULT 2

US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782

Query Match 10.8%; Score 111; DB 10; Length 6728;
Best Local Similarity 53.0%; Pred. No. 3.7e-20;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 122 TACAGGTTCACCGGGCCCCAGGGTCCAGAGAGGTGACAGAGATCCCAAGGACCCCTG 181
DB 721 TCCAAAGTCCCCCTGGTGAGCTGGCGAGCCCTTCAGGTCCCAATGGTCCCGGAG 780
QY 182 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGGGGAGCCCTGACACCACTGGCCCTG 241
DB 781 GTCCCCCAGGTCCCCCTGGAAAGATGAGATGATGGGGAAGCTGGAACACCTGTCGTC 840
QY 242 CGGGTGAAGAGGCCCAATTTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGAT 301
DB 841 CTGTGTGAGCGTGGGCTCTCTGGGCTCAGGTGCTCGAGGATTTGCCCGGAACAGCTGGCC 900
QY 302 CTAAGAGCTCCAGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGG 361
DB 901 TCCCTGGAATGAAGGACACAGAGGTTTCAGTGGTTTGGATGTTGCCAAGGAGATGCTG 960
QY 362 GCCCCAGTGGGACCCAGGCCCCCGGGCCCAAGAGGAGGACTCCCGGCGCCCTC 421
DB 961 GTCCCTGCTGGTCTTAAGGTTGAGCCTGGCAGCCCTTGGTGAATGAGCTCCTGCTCAGA 1020

QY	422	AGGGCCCTCTGCTTCAGAGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGTGCTGGAC	481
Db	1021	TGGGCCCCCGTGGCCCTGCGCTGTGAGAGAGGTGCGCCCTGGAGCCCTGGCCCTGCTGTGTG	1080
QY	482	CTGGGGGACTGCCAGCGTTGCGCTGGGTGACCAAGGCATGCCAGGGCCCCCAAGGGCCCCCCCCG	541
Db	1081	CTGCTGGAATGATGTGTCTACTGTGTGTCGGGGCCCCCTGGTCCCAACCGGGCCCCGCTG	1140
QY	542	GCCTCTCTGGCCCATCAGAGGCGGTGG	568
Db	1141	GTCTCTGCGTTCCCTGTGTGTGTG	1167

RESULT 3

US-09-880-107-3946

: Sequence 3946, Application US/09880107

: Patent No. US20020142981A1

: GENERAL INFORMATION:

: APPLICANT: Horne, Darci T.

: APPLICANT: Vockley, Joseph G.

: APPLICANT: Scherf, Uwe

: APPLICANT: Gene Logic, Inc.

: TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

: FILE REFERENCE: 44921-5028-WO

: CURRENT APPLICATION NUMBER: US/09/880,107

: CURRENT FILING DATE: 2001-06-14

: PRIOR APPLICATION NUMBER: US 60/211,379

: PRIOR FILING DATE: 2000-06-14

: PRIOR APPLICATION NUMBER: US 60/237,054

: PRIOR FILING DATE: 2000-10-02

: NUMBER OF SEQ ID NOS: 3950

: SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 3946

: LENGTH: 6728

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z74615

US-09-880-107-3946

```

Db      1141   GTCTCCTGGCTTCCCTGGTGCTGTG    1167
          | ||||| ||| | || || || |
RESULT 4
US-09-924-340-57
; Sequence 57, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US.2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
US-09-924-340-57

```

Query Match	10.8%;	Score 111;	DB 10;	Length 6728;
Best Local Similarity	53.0%;	Pred. No. 3,7e+0;		
Matches 237;	Conservative 0;	Mismatches 210;	Indels 0;	Gaps 0;
QY	122	TACAAGGTCCACCGGGCCCAAGGGGTCCAAGAGGTGCACAGAGATCCCCAGGAGCCCCCTG	181	
DB	721	TCCAAGTCCCCCTGGTGAGCCTCGGAGCCTGGAGCTTCAGTGCCTATGGTCCCCGAG	780	
QY	182	GCCCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGCCTGGACCACTGSCCCTG	241	
DB	781	GTCCCCCAGGTCCCCCTGGAAAGAAATGAGAGATGATGGGGAAGCTGGAAAACCTGTCGTC	840	
QY	242	CGGTGTGAGAGAGGCCAATTCGGACAGCTGTCCCCCGGAGAGCGTGCSCGCAAAAGAT	301	
DB	841	CTGTTGAGCTGGGCCTCTGGGCTCTCAGGCTGCTCAGGATGTCGCCGGAACAGCTGGCC	900	
QY	302	CTAAAGGCTCCAGGGGCCCAAGAGCTCCCGTGGTTCCCTCGGAAGCCCGCCCTCAGG	361	
DB	901	TCCCTGGAATGAAGGGACACAGAGGTTTCAGTGGTTTGGATGGTCCCAAGGAGATGCTG	960	
QY	362	GCCCCAGTGGGACCCAGGCCCCCGGGCCACACAGAGCAAGAGGAGCTCCCCGGCCCTC	421	
DB	961	GTCTGTGTGTCCTAAGGGTGAGCTGTGCAGCCCTGGTGGAAATGGAGCTCCTGGTCA	1020	
QY	422	AGGGCCCTCTCGGTTCCAGGGAGCTTCAGGCAACCGTTTGGGGAGCCTGGGGTGCCTGGAC	481	
DB	1021	TGGGCCCCCGTGGCCTGCCTGGTGAGAGAGTGCGCCCTGGAGCCCTGGCCCTGCTG	1080	
QY	482	CTCGGGGACTCCAGGCTTGGCTGGGTACAGAGCATCCAGGCCCCCAAGGGCCCCCCCCG	541	
DB	1081	CTCTGGAAATGATGTGTGCTACTGTGTGTCGGGGGCCCTCTGGTGTCCACCGGCCCGCTG	1140	
QY	542	GCCCTCTCTGGCCCCATCAGGAGCGGTGG	568	

```
QY 501 GCCTGGGTACAGGATCCAGGCCCCCAAGAGCCCCCGCCCTCTCTGGGCCCATCAGG 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TAAGGTGACTGTGGCAACACAGGTCTCTGTGCAGCACTGGCGGCTGGCGCAGAGG 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 AGCGGTGGTGGCCCTGGCCCTGCAG 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 TGAACCTGGTGCATGGGACCCAG 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.054.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 476..964
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
US-09-992-600A-57

Query Match 10.8%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 1.8e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

QY 90 GCATGGTCAGCTCAACAGATTTTACAACTACTACAAGTCCACCGGGCCCCAGGGTCC 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 GCAAGAGCAGGGCCCCCAGGAGTCTGAAGGCTCTCCAGGGCCCCCAGGGCCCCAGGGTCC 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 150 AAGAGGTGACAGAGGATCCAGGGGACCCCTGGCCCACTGGCAACAAGGGACACAAAGG 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 CAGAGAGCAGGACGACCCACAGGTAACTCCGGTGAGAGGCGCACAGGAGTTTCAAGG 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 210 AGAAGAGGGGAGCTGGACCACTGGCCCTGCGGGTGAGAGAGGCCCAATTTGGACCAGC 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 CCAGCAGGCTTTTCGGGGCCACCGGGTCCCTCGATTCCCGACCAAAAGTTGGATCACC 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 270 TGGTCCCCCGG-----AGACCGTGGCGGCAAGATCTAAAGGCTCCACAGGGCCC 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 289 TGGCCCACTGCCCCCTCAAGCAGAGAAAGGCGACGAAAGGATTCAGAGGGCCCCATCAGGCCT 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 321 CAAAGGCTCCCTGCTGTCCTGGGAAGCCCGGCTCTCAGGGCCCCAGTGGGGACCCAGG 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 GCCTGGTCCCTGCGCCACCGGGACCTCTGGGATTACAGGCCCCCGCGGTCTGGATGG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 381 CCCCCCGGCCCCACCAAGAGGAGGACTCCCGGCCCCCTCAGGGCCCCCTCTCTGGCTTCCA 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 TTTGGATGGGAAGGATGGCAAGCCTGGCTTAGGGGGACCCCTGTCTCTGCTGGCCCCC 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 441 GGGACTTCAGGCACCGTTGGGGAGCCTGGGGTGGCTGGACCTCGGGGACTGCCAGGCTT 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 TGGACTCATGGGACCAACCGGCTTTAAGGGGAAACAGGACATCTCTGGCTCCCAAGACC 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 501 GCCTGGGTACCAAGGCATGCCAGGCCCAAGGGCCCCCGGCCCCCTCTCTGGCCCATCAG 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 TAAGGTGACTGTGCAACACAGGTCTCTCTGGCAGCACTGGCCGCTTGGCGCAGAGG 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 AGCGGTGGTGGCCCTGGCCCTGCAG 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 TGAACCTGGTGCATGGGACCCAG 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-09-925-302-64
; Sequence 64, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 2691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2653)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2667)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2683)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-64

Query Match 10.7%; Score 109.4; DB 10; Length 2691;
Best Local Similarity 54.4%; Pred. No. 6.2e-20;
Matches 240; Conservative 1; Mismatches 197; Indels 3; Gaps 1;

QY 124 CAAGTCCACGGGGCCCCCAGGGTCCACAGAGGTGACAGAGGATCCAGGGACCCCTGGC 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 CCAGGACACCGGTTGTCCAGGGCCCCAAGGGAGACCCCTGGATTCCAGGGCATTCGCTGT 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 CCAACTGCCAACACAGGACAGAGAGGGAGAGAGGGGAGCCCTGGACCACTCGCCCTGG 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 ATTGTGGCTCTCCAGGAATCACAGGTCTTAAGGTGATATGGGGCTCCAGAGAGTTCCA 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 GGTGAGAGAGGCCCAATTGGACCACTGTCTCCCGCGGAGAGCGCTGGGGCAAGGATCT 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 GGATTTCAAGGTCCAAAGAGTCTTCTCGCCCTCCAGGGAATTAAGGTGATCAAGGCGAT 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 AAAGCTCCCGGGCCCCCAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAGG-- 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db	211	CAWGGCTCCCGGGAGCTAAAGTCTCCGGGTCTCTTGGCCCCCAGGTCCTTACGAC	270
Qy	362	-GCCCAAGTGGGACCCAGGCCCCCCCGGGCCCAACGAGCAAAAGAGGACTCTCCCGGCCCT	420
Db	271	ATCATCAAGGGAGGCCCGGGCTCCCTGTGTCTTGGGGCCCCCAGGGCTGAAGGGGTT	330
Qy	421	CAGGGCCCTCTGGCTTTCAGGACATTCAAGGCACCGTTGGGAGCCTGGGTGCTTGA	480
Db	331	CAGGACTGCCAGGCCCGAAAGCCAGCAAGGTGTTACAGGATTGGTGGTATACCTGA	390
Qy	481	CCTCGGGGACTGCCAGGCTTGCCTGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCCC	540
Db	391	CCTCCAGGTATCTGGTTTGAGGTGCCCTGGCCAGAAAGGAGAGATGGGACCTGCC	450
Qy	541	GGCCCTCTGGCCCATCAGA	561
Db	451	GGGCTACTGTCCAAAGGA	471

```

RESULT 7
US-09-925-301-42
; Sequence 42, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925 301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; . LENGTH: 2192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-42

```

Query Match	10.4%	Score 107.2	DB 10	Length 2192
Best Local Similarity	53.5%	Pred. No. 2.2e-19		
Matches 274	Conservative 0	Mismatches 228	Indels 10	Gaps 2
QY 127	GGTCACCGGCCCCAGGGGTCCAAAGAGTGCACAGAGATCCACGAGACCCCTGGGCCA	186		
Db				
QY 638	GGCCCCGTGGACCCCTGGTCTCTGGTGCTCTCTGGTCCCTTGGCCCCGTGGCCCT	697		
Db				
QY 187	ACTGGCAACAGGACAGAAAGAGAGAGGGGAGCTGGACACCTTGGCCCTTCGGSGT	246		
Db				
QY 698	GCTGGCAAGAGTGTGATCGTGGTGAGACTGGTCTCTGGTCCCGCGTCTGTGCGC	757		
Db				
QY 247	GAGAGAGGCCAATTGACACAGCTGCTCCCGCGAGAGCGTGGCGGCAAGGATCTAAA	306		
Db				
QY 758	CTCTTTGGCCCGTGGCCCGCGCGACCCCAAGGCCCTGGTGACAAAGGGTGAGACA	817		
Db				
QY 307	GGCTCCACGGGCCCAAGGCTCCCGTGTCTCCCTGGGAAGCCCGGCCCTTCAGSGGCC	366		
Db				
QY 818	GGCAACAGGCGACAGAGCATAAAGGGTCACCGTGGCTTCTCTGGCCTCCAGSGTCC	876		
Db				
QY 367	AGTGGGACCCAGGCCCGGGGCCACCAAGGAAAGAGGACTCCCGGGCCCTCAGGGC	426		
Db				
QY 877	-----CCCTGGCCCTCTGGCTCTCTGGTGAACAAGTCCCTCTGGAGCCTCTGCT	928		
Db				
QY 427	CCTCTGGCTTCCAGGACTTCAGGGCACCGTCTGGGAGCCTTGGGTCCTTGGACCTCGG	486		
Db				
QY 929	CCTGCTGGTCCCGAGGTCCTCCCTGGCTCTGCTGCTCTCTGGCAAGATGGAATCAAC	988		
Db				
QY 487	GGACTGCCAGCTTGCTGGGTACAGGCATGCCAGGCCCGCCCGGCCCTTCCGGCCCT	546		
Db				
QY 989	GGTCTCCCTGGGCCCAATGGGCCCTTGCTCTCGCGTGCACCTGGTGTGATCTGGTCT	1048		
Db				

Qy	547	CCTGGCCCATCAGAGCGG-TGGTGGCCCTTGCGCCCTGCAGAAATGACCAACCCCGCACCC	605
Db	1049	GTTGTGCCCCCGGCCCTCTCGACCTCTCGTGGTCGCCCTGGTCTCCACAGCGCTGGTTTC	1108
Qy	606	GGAGGACAATGGCTGGCCCGCTCAGCTGGRAGA	637
Db	1109	GACTTCAGCTCTCTGCCCCAGCCACTCAAGA	1140
 RESULT 8			
US-09-880-107-3947			
; Sequence 3947, Application US/09880107			
; Patent No. US20020142981A1			
; GENERAL INFORMATION:			
; APPLICANT: Horne, Darci T.			
; APPLICANT: Vockley, Joseph G.			
; APPLICANT: Scherf, Uwe			
; APPLICANT: Gene Logic, Inc.			
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer			
; FILE REFERENCE: 44921-5028-WO			
; CURRENT APPLICATION NUMBER: US/09/880,107			
; CURRENT FILING DATE: 2001-06-14			
; PRIOR APPLICATION NUMBER: US 60/211,379			
; PRIOR FILING DATE: 2000-06-14			
; PRIOR APPLICATION NUMBER: US 60/237,054			
; PRIOR FILING DATE: 2000-10-02			
; NUMBER OF SEQ ID NOS: 3950			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 3947			
; LENGTH: 5086			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616			
; NAME/KEY: unsure			
; LOCATION: (1)..(5086)			
; OTHER INFORMATION: n = a or c or g or t			
US-09-880-107-3947			

Query Match	10.4%;	Score 107;	DB 10;	Length 5086;
Best Local Similarity	51.3%;	Pred. No. 4e-19;		
Matches 248;	Conservative 0;	Mismatches 235;	Indels 0;	Gaps 0;
Qy	120	ACTAAGGTCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCC	179	
Db	709	ACAGCCCGTCTCTGCTGTAAGGTTGAACCTGGTGCCTCGTGAATGAAATGCAACTCC	768	
Qy	180	TGGCCCCAACTGGCAACAGAGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCTCTGGCC	239	
Db	769	AGGTCAAAACAGAGCCCGTGGGCTTCCTGGTGAAGAGAGACGTGTGGTGGCCCTGGCC	828	
Qy	240	TGCGGTTGAGAGAGGCCCAATTTGACACAGCTGGTCCCCCGAGAGCGTGGCGCAAGG	299	
Db	829	AGCTGGTGCCCGTGGCAGTGAATGGAAGTGGGTGCCGTGGGTCTCTGTTGCTCCATTGG	888	
Qy	300	ATCTAAAGGCTCCACAGGGCCCCAAAGGCTCCCGTGGTTTCCCTGGGAAGCCCGCCCTCA	359	
Db	889	GTCTGCTGGCCCTCCAGGCTTCCAGCTGCCCTGGCCCCCAAGGTTGAAATGGAGTG	948	
Qy	360	GGGCCCCAGTGGGAGCCAGGCCCCCGGGCCCCACCAAGCAAAAGAGGACTTCCCGGGCC	419	
Db	949	TGGTAAGCGTGGTCTGCTGGTCCCGGGTCCCGGTGGTGAAGTGGGTCTTCCAGGCGCT	1008	
Qy	420	TCAGGGCCCTCTGGCTTCCAGGAGCTTCAGGGGACACGGTTGGGAGGCTGGGGTGCCTGG	479	
Db	1009	CTCGGGCCCCGTTGAGACTCTGCTAATCTTGAGACAAACGGCCCTTACTGGTGCAAGG	1068	
Qy	480	ACCTGGGGACTGCCAGGCTTGCCTGGGTACAGGATGCCAGGCCCAAGGGCCCCC	539	
Db	1069	TGCTGCTGGCCCTTCCCGCGTGTGTGGGCTTCCCGGCTCCCTGGACCCCGCGGTATTC	1128	
Qy	540	CGGCGCTCTCTGGCCCATCAGGAGCGGTGTGGCCCTGGCCCTGCAGAAATGAGCAACCC	599	

Db 1129 TGGCCCTGTGTGGTGGCGGTGCTACTGTGCCAGAGACTTGTGTGAGCCTGGTCC 1188
QY 600 GGC 602
Db 1189 AGC 1191

RESULT 9
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 10.4%; Score 107; DB 9; Length 5145;
Best Local Similarity 51.3%; Pred. No. 4e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;
QY 120 ACTACAAGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGATCCCGAGGGACCC 179
Db 712 ACAGCCGGTGTCTGTGGTGAAGGGTGAACCTGTGTGCCCTTGGTGAAGTGAAGTGA 771
QY 180 TGGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGCCCTGGACCACTGGCC 239
Db 772 AGGTCAACAGAGCCGTGGGCTTCTGTGTGAGAGAGACGTGTGTGGTCCCGCTGGCC 831
QY 240 TGGCGGTGAGAGAGGCCCAATTGACACGTGTGTCCCGCCGAGAGAGCGTGGCGGCA 299
Db 832 AGCTGTGTCGCTGGCAGTGTGAAGTGTGGGTCCCGTGGGTCTGTGTGGTCCCAATTGG 891
QY 300 ATCTAAGGCTCCAGGGCCCAAGGCTCCCGTGTGTTCCTTGGCAAGCCCGGCCCTCA 359
Db 892 GTCTGTGTCGCTGACAGGTTCACAGGTTCGCCCTGGCCCAAGGGTGAATTTGGARCTGT 951
QY 360 GGGCCCACTGGGGACCCAGGGCCCGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
Db 952 TGGTACAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1011
QY 420 TCAGGCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 479
Db 1012 CTCGCGCCCGCTGTGACCTTCCTGTTAATCTGTGAGCAACGGCCTTACTTGGTGCCA 1071
QY 480 ACCTCGGGGACTCCAGGCTTGCCTGGGTACCGACATGCCAGGCCCAAGGGCCCCCCC 539

Db 1072 TCCTGTGTGGCTTCCCGGGTGTGTGGGGTCCCGGGCTCCCTGGACCCCGGGTATTCC 1131
QY 540 CGGCGCTCTGTGGCCCATCAGGAGCGGTGTGTGGCCCTGTGCGCTGGCCCTGACAGATGAGCCACCC 599
Db 1132 TGGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1191
QY 600 GGC 602
Db 1192 AGC 1194

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 10.4%; Score 107; DB 10; Length 5145;
Best Local Similarity 51.3%; Pred. No. 4e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;
QY 120 ACTACAAGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGATCCCGAGGGACCC 179
Db 712 ACAGCCGGTGTCTGTGGTGAAGGGTGAACCTGTGTGCCCTTGGTGAAGTGAAGTGA 771
QY 180 TGGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGCCCTGGACCACTGGCC 239
Db 772 AGGTCAACAGAGCCGTGGGCTTCTGTGTGAGAGAGACGTGTGTGGTCCCGCTGGCC 831
QY 240 TGGCGGTGAGAGAGGCCCAATTGACACGTGTGTCCCGCCGAGAGAGCGTGGCGGCA 299
Db 832 AGCTGTGTCGCTGACAGGTTCACAGGTTCGCCCTGGCCCAAGGGTGAATTTGGARCTGT 891
QY 300 ATCTAAGGCTCCAGGGCCCAAGGCTCCCGTGTGTTCCTTGGCAAGCCCGGCCCTCA 359
Db 892 GTCTGTGTCGCTGACAGGTTCACAGGTTCGCCCTGGCCCAAGGGTGAATTTGGARCTGT 951
QY 360 GGGCCCACTGGGGACCCAGGGCCCGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
Db 952 TGGTACAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1011
QY 420 TCAGGCGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 479
Db 1012 CTCGCGCCCGCTGTGACCTTCCTGTTAATCTGTGAGCAACGGCCTTACTTGGTGCCA 1071


```

; Sequence 961, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 961
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-961

Query Match          10.3%; Score 106; DB 9; Length 2542;
Best Local Similarity 52.1%; Pred. No. 5.1e-19;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 122 TACAAGTCCACGGGCCCGCCAGAGAGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCTCG 181
Db 630 TCCTTGACACCGGTGCCAGGCGATCAGGGCTACCCAGGCATGGCAGGCGCCCAAGG 689
QY 182 GCCCACTGGCAACAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
Db 690 GAGAGACGGGCGCTCATGGATATAAAGGCATGTTGGGCGCTATCGGTGCCACTTGGGCCAC 749
QY 242 CGGGTCCAGAGGCGCCCAATTTGACAGAGTCCCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 301
Db 750 CGGGTCCAGAGGCGCCCAATTTGACAGAGTCCCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 809
QY 302 CTAAAGGTCCTCCAGGCGCCCAAGAGGCTCCCGTGGTTCCCTGGGAAGAGAGAGAGAGAGAGAG 361
Db 810 GCCAGGTATTCGTGGAGCCCGGAGGATCACAGGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 869
QY 362 GCCCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
Db 870 GCATCAACGGCAAGGATGGGACCGCCAGGCGATGAGGCGCTATCGGTGCCACTTGGGCCAC 929
QY 422 AGGGCCCTCTGCTTCCAGGAGCTTCCAGGAGCTTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
Db 930 CGGGACAGCCCGGAGAGTCCAGGCGCCAGGCGCTAGCGGGTGTGCCAGGCGAGCTTGGGA 989
QY 482 CTGCGGAGTCCAGGCTTCCCTGGGTTACAGGATGACAGGCGCCCAAGAGAGAGAGAGAGAGAGAG 541
Db 990 CAAAAGGAGCGCTTGAGAGACAGGAGTGGAGCGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
QY 542 GCCCTCTGCGCCATCAGAGCGGTGGTCCCG 574
Db 1050 GTCCCCCTGGGAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082

RESULT 14
US-09-964-824A-255
; Sequence 255, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrihan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-73

```

```

; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 255
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-255

Query Match          10.3%; Score 106; DB 10; Length 2542;
Best Local Similarity 52.1%; Pred. No. 5.1e-19;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 122 TACAAGTCCACGGGCCCGCCAGGAGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCTCG 181
Db 630 TCCTTGACACCGGTGCCAGGCGATCAGGGCTACCCAGGCATGGCAGGCGCCCAAGG 689
QY 182 GCCCACTGGCAACAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
Db 690 GAGAGACGGGCGCTCATGGATATAAAGGCATGTTGGGCGCTATCGGTGCCACTTGGGCCAC 749
QY 242 CGGGTCCAGAGGCGCCCAATTTGACAGAGTCCCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 301
Db 750 CGGGTCCAGAGGCGCCCAATTTGACAGAGTCCCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 809
QY 302 CTAAAGGTCCTCCAGGCGCCCAAGAGGCTCCCGTGGTTCCCTGGGAAGAGAGAGAGAGAGAGAG 361
Db 810 GCCAGGTATTCGTGGAGCCCGGAGGATCACAGGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 869
QY 362 GCCCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
Db 870 GCATCAACGGCAAGGATGGGACCGCCAGGCGATGAGGCGCTATCGGTGCCACTTGGGCCAC 929
QY 422 AGGGCCCTCTGCTTCCAGGAGCTTCCAGGAGCTTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
Db 930 CGGGACAGCCCGGAGAGTCCAGGCGCCAGGCGCTAGCGGGTGTGCCAGGCGAGCTTGGGA 989
QY 482 CTGCGGAGTCCAGGCTTCCCTGGGTTACAGGATGACAGGCGCCCAAGAGAGAGAGAGAGAGAGAG 541
Db 990 CAAAAGGAGCGCTTGAGAGACAGGAGTGGAGCGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
QY 542 GCCCTCTGCGCCATCAGAGCGGTGGTCCCG 574
Db 1050 GTCCCCCTGGGAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082

RESULT 15
US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637

```



```
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 786
; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-786

Query Match      10.3%; Score 105.4; DB 10; Length 5416;
Best Local Similarity 51.1%; Pred. No. 1.1e-18;
Matches 247; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 120 ACTACAAGGTCACCGGCCCCAGGGGTCACAGAGGTACACAGAGATCCAGGGACCCCC 179
Db 1039 ACAGCCCGGTCTCTGTGTGAAGGTGAACCTGGTGCCTCTGTAATAATGGAACCTCC 1098

QY 180 TGGCCCAACTGGAACAAGGACACAAAGAGAGAGAGGGGAGCCTGGACCACTGGGCC 239
Db 1099 AGGTCAACAGAGAGCCCTGGTCTTCCTGGTGAGAGAGAGCGTGTGGTGCCTCTGGTCC 1158

QY 240 TCGGGGTGAGAGAGCCCAATTGGACCACTGGTCCCCCGAGAGAGCGTGGCGGCAAGG 299
Db 1159 AGCTGGTCCCGTGAAGTGTGAAGTGTGGTCCCGTAGTCTCTGGTCTCTAATGG 1218

QY 300 ATCTAAAGGCTCCAGGCCCCCAAGAGCTCCGCTGGTTCCTGGGAAGCCCGGCCCTCA 359
Db 1219 GTCTGCTGGCCCTCCAGGTTTCCAGAGTGCCTCTGTCCTCCCAAGGTTGAATTTGAGCTGT 1278

QY 360 GGGCCCCAGTGGGGACCCAGGCCCCCGGCCACCAGGCAAGAGGACTCCCGGGCCC 419
Db 1279 TGGTAACGCTGGTCTACTTGGACCCCGCGTCCCGTGGTGAAGTGGGTCTTCCAGGCT 1338

QY 420 TCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGCCTGG 479
Db 1339 CTCGGCCCCGTTGGACCTCCTGGTAATCCTGGAGCAAGGSCCTTACTGGTGCCAGGG 1398

QY 480 ACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGATGCCAGGCCCCAGGGGCCCCCC 539
Db 1399 TGCTGCTGGCCCTTCCCGCGTGTCTGGGGCTCCCGGCTCCCTGGACCCCGCGGTATTCC 1458

QY 540 CGGCCCTCTGGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAGATGAGCCCAACCCC 599
Db 1459 TGGCCCTCTGGTGTGCCGTACTACTGTGTCACAGAGACTTGTGGTGAAGCTGGTCC 1518

QY 600 GGC 602
Db 1519 AGC 1521
```

Search completed: March 21, 2003, 08:22:40
Job time : 175.339 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 1346.02 Seconds
(without alignments)
12344.923 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_1695
Perfect score: 1026
Sequence: 1 atcaacaagaattgatgag.....cagtactgtcatctgcatta 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estov:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	800.4	78.0	3305	11	BC009162 Mus muscu
2	556.4	54.2	861	14	B0713873 AGENCOURT
3	525.4	51.2	668	14	B0674807 AGENCOURT
4	475.4	46.3	967	9	AL568743
5	462.6	45.1	906	13	B1456109
6	445.6	43.4	580	12	BE910803

7	437.2	42.6	552	10	AW958053
8	415.8	40.5	601	10	BE290299
9	403.4	39.3	957	14	BQ955927
10	369.4	36.0	638	10	BB248064
11	351.4	34.2	500	14	BM676508
12	350	34.1	507	14	BM713891
13	343.8	33.5	808	14	BQ771366
14	311	30.3	339	9	AA304251
15	301.4	29.4	357	10	AW379436
16	297.2	29.0	517	9	AL543000
17	297	28.9	936	14	BQ891432
18	296	28.8	310	9	AA361740
19	271	26.4	350	9	AA012704
20	245.8	24.0	420	12	BE929813
21	233.8	22.8	429	12	BE929798
22	212	20.7	723	14	BQ004176
23	210.2	20.5	943	14	BQ887163
24	201	19.6	655	12	BG482931
25	200.8	19.5	345	9	AI7742661
26	199.6	19.5	441	10	BE749743
27	185.2	18.1	591	13	BI445884
28	185	18.0	277	12	BF854320
29	183.4	17.9	234	12	BF854188
30	183.4	17.9	261	13	BI010880
31	180.2	17.6	280	13	BI010883
32	178.8	17.4	286	12	BF854324
33	172.4	16.8	334	10	AW002141
34	170.4	16.6	309	14	W76214
35	159.8	15.6	281	12	BF854408
36	151.2	14.7	465	10	AW240221
37	144.4	14.1	504	14	R74387
38	142.8	13.9	288	14	W72977
39	135.6	13.2	432	9	AI613616
40	127.2	12.4	1082	14	BM907108
41	124.4	12.1	884	14	BQ934501
42	124.2	12.1	531	10	AV603727
43	123.8	12.1	608	10	AW918393
44	122.4	11.9	960	14	BQ920508
45	118.8	11.6	824	13	BI828870

ALIGNMENTS

RESULT 1	BC009162	BC009162	3305 bp	mRNA	linear	HTC	07-AUG-2002
LOCUS	BC009162	Mus musculus, Similar to scavenger receptor with C-type lectin, clone IMAGE:2811487, mRNA.					
DEFINITION	BC009162	BC009162.1	GI:14714370				
ACCESSION	BC009162	HTC					
VERSION	BC009162.1	house mouse.					
KEYWORDS	HTC	Mus musculus					
SOURCE	BC009162.1	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
ORGANISM	BC009162.1	1 (bases 1 to 3305)					
REFERENCE	BC009162.1	Strausberg, R.					
AUTHORS	BC009162.1	Direct Submission					
TITLE	BC009162.1	Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA					
JOURNAL	BC009162.1	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabbs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/					
REMARK	BC009162.1						
COMMENT	BC009162.1						

Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 5 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

source

Location/Qualifiers

1..3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="mammary tumor. WAP-rtgF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"

BASE COUNT 1031 a 757 c 760 g 757 t

ORIGIN

Query Match 78.0%; Score 800.4; DB 11; Length 3305;
Best Local Similarity 86.3%; Pred. No. 2.4e-179;
Matches 885; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGCAACAAGATTGATGAGGTGCGAGGTGACACTGAAGTAGCCCAACTTATCAGTGATT 60
DB 1296 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTAGCCCAACTTATCAGTGATT 1355
QY 61 ATGGAAGAATAGCTAGTACGCTCCACAGCATGCTCAGCTCATCAAGAAATTTTACATA 120
DB 1356 ATGGAAGATGAATCTGTTGACTCCAAAGCAGCGGTGAGCTCATCAAGAACTTTTACCA 1415
QY 121 CTAAAGGTCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 180
DB 1416 CTAAAGGTCTCTCTGGCCCCAGAGGTCCAAAGGTGACAGAGGATCTCAGGACCACT 1475
QY 181 GCGCCCACTGGCAACAGAGGACAGAAAGAGAGAGGGGAGCCCTGGACCACTTGGCCCT 240
DB 1476 GTTCAACTTGGCAACAGAGGACAGAAAGAGAGAGGGGAGAGCCCTGGTCCACTGGCCCT 1535
QY 241 GCGGGTGAGAGAGCCCAATTTGACACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGA 300
DB 1536 GCGGGTGAGAGGGGCACAAATTTGACACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGA 1595
QY 301 TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 360
DB 1596 TCCAAGGCTCAGGGTCCCAAGAGATCTCGTGGGTCCCGAGGAAAGCGTGGCCCTCAA 1655
QY 361 GCGCCCACTGGGACCCAGGCCCCCGGGCCCCACACAGGCAAGAGAGGACTCCCGGCCCT 420
DB 1656 GGACCTAGTGGGGAGCCAGGACCAAGGTCACAGGCAAGGAGTGGACTCCCTGGCCCT 1715
QY 421 CAGGCCCCCTCTGGCTTCCAGGACTTCCAGGACCGTTGGGGAGCGTGGCGTGGCTTGA 480
DB 1716 CAGGCCCCCTCTGGCTTCCAGGACTTACAGGCACTGTGGGTGAGCCTGGAGTACTTGA 1775
QY 481 CCGGGGACTCCAGGCTTGGCTGGGTACCAGGCTGCCAGGCCCCCAAGGCCCCCCCC 540
DB 1776 CCGGGGGTTTCCAGGCTTGGAGGGGTGCCAGGATGCCCTGGGCTTAAGGACCACT 1835
QY 541 GCGCCCTCTGGCCATCAGGAGCGGTGTGGCCCTTGGCCCTGCAGAAAGAGGACCCCG 600
DB 1836 GCGCCCTCCAGGCCCCCTCAGAGCAATGGAGCAATTTGGCTCTGCAGAAATGAACCAACCCA 1895
QY 601 GCACGGAGCAATGGCTGCGCGCTCACTGGAAGACTTCCACAGACAAATGCTACTAT 660
DB 1896 GCATAGAGGTCAAGGATGTCCACTCTAGGAAGACTTCCACAGATAAATGCTACTAT 1955

QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGTAAGACAAGTCTTCA 720
DB 1956 TTTTCATTGGAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGTAAGACAATCTTCC 2015
QY 721 CATCTTTGTTTCATTAACACTAGAGAGGAACAGCAATGCAAAAAACAGATGGTAGGG 780
DB 2016 CATCTCGTTTTCATTAACACTCAAGAGAAAGAACGCAATGATGATAAAAGCATACCGTGGG 2075
QY 781 AGAGAGAGGCACTGGATCGGCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 840
DB 2076 AGAGAAAGCCATGGATCGGCTCACAGACTCAGAACAGAAAGCAATGGAAGTGGCTA 2135
QY 841 GATGGACATCTCCAGACTTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTCTAT 900
DB 2136 GACGGTCACTGTTTCATTACAAAACCTGAAAGCTGGACACCAAGATAACTGGGCACT 2195
QY 901 GGCATGGCCAGGAGAGAGCTGTGCTGGTTGATTATGCTGGGAGTGGAGCAATTC 960
DB 2196 GGCATGGCCAGGAGAGAGCTGTGCTGGTTGATTATGCTGGGAGTGGAGCAATTC 2255
QY 961 CAATGTGAAGACGTCAATAACTTCTTCGAAAAAGACAGGAGACACTACTGTCTCATCT 1020
DB 2256 CAGTGTGATGAATCAATAACTTCTTGTGAGAAAGGAGGACGACATCATCTCC 2315
QY 1021 GCATTA 1026
DB 2316 ATATTA 2321

RESULT 2

BQ713873

LOCUS

BQ713873 861 bp mRNA linear EST 16-JUL-2002
AGENCY: 8482828 NIH_MGC_129 Mus musculus cdna clone IMAGE:6306240
5', mRNA sequence.

ACCESSION

BQ713873

VERSION

BQ713873.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 861)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

cDNA Library Preparation: ResGen, Invitrogen Corp

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13721 row: n column: 01

High quality sequence stop: 598.

FEATURES

source

1..861

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:6306240"

/clone_lib="NIH_MGC_129"

/lab_host="DH10B (phage-resistant)"

/note="Organ: olfactory epithelium; Vector:

PCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned

unidirectionally. Primer: Oligo dr. Average insert size

2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this

is a NIH_MGC Library."

BASE COUNT 221 a 253 c 239 g 147 t 1 others

ORIGIN

Query Match 54.2%; Score 556.4; DB 14; Length 861;

Best Local Similarity 82.8%; Pred. No. 1.4e-121; Matches 647; Conservative 0; Mismatches 132; Indels 2; Gaps 1;			
Qy	14	TGATGAGTCCAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATATGGAAGAAATGA	73
Db	14	TGATGAGTCAAGATTAGACACTGAAGTGGCCAACTTATCAGTGGTTATGGAAGATGA	73
Qy	74	AGCTAGTAGATCCCAAGCATGGTCAGCTCATCAAGATTTTACAAATCTACAAGGTCCAC	133
Db	74	AACGTGTGACTCCCAAGCACGGTCAGCTCATCAAGACTTTTACCATTCTACAAGGTCTC	133
Qy	134	CGGGCCCCAGGGTCCAGAGAGGTGACAGAGATCCCAAGGACCCCTGGCCCCCACTGCCA	193
Db	134	CTGGCCCCAGAGGTTCCAAAGGTGACAGAGATCTCAGGGACCACTGGTCCCACTGCCA	193
Qy	194	ACAAGGACACAGAAAGGAGAGAGGGGAGCTTGGACCACTGGGCCCTCGGGTGGAGAG	253
Db	194	ACAAGGACACAGAAAGGAGAGAGGGAGAGCTTGTCCACCTGGGCCCTCGGGTGGAGAG	253
Qy	254	GCCCAATTGGACAGCTGCTCCCGGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCC	313
Db	254	GCACAAATTGGACAGCTGCGGCCCTCTGGAGAGCGTGGCAGCAAGAGATCCAAAGGCTC	313
Qy	314	AGGGCCCCAAGAGCTCCCGTGGTCCCTGGGAGAGCCCGGCCCTCAGGGCCCCAGTGGG	373
Db	314	AGGGTCCCAAGAGATCTCGTGGGTCCCGAGGAGAGCCCTGGGCCCTCAAGGACCTAGTGGG	373
Qy	374	ACCAGGGCCCCCGGGCCACAGCAAGAGGAGTCCCGGCCCTCAGGGCCCCCTCCTG	433
Db	374	ACCAGGACCAAGGATCCACAGCAAGAGATGAGTCCCTGGGCCCTCAGGGCCCCCTG	433
Qy	434	GCTTCCAGGGACTTCAGGGCACCGTGGGAGCGCTGGGGTGCCTGGACCTCGGGGACTGC	493
Db	434	GCTTCCAGGGACTTCAGGGCACCTGTGGGTGAGCCTGGAGTACCTGGACCTCGGGGTTGC	493
Qy	494	CAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCTCCTG	553
Db	494	CAGGCTTGCCTGGGTACCAAGCATGCCAGGGCCCTGGGGCCCTTAAGGGACCACTGGGCCCTCAGGGCC	553
Qy	554	CATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAAGCCCGGCACCGGAGACA	613
Db	554	CCTCAGGAGCAATGGAGCCATTGGCTGTGCAAAATGAAACCAACCCCAACATCAGAGTCA	613
Qy	614	ATGCTGCCCCCTCTACTGGAAGAACTTCACAGACAAATGTACTATTTTTCAGTTGAGA	673
Db	614	ACGGATGTCGCCCTCACTGGGAAGAACTTCACAGATAAATGTACTATTTTTCCTTGGCAA	673
Qy	674	AAGAAATTTTGGAGTCAAGC--TTTTCTGTGAAGACAAGTCTTCACATCTTGTTT	731
Db	674	AAAAAATTTTGCAGAAATGTACCTTTTCTGCGAACAATAATCTCCCCCACTCCGGGTTT	733
Qy	732	CATAAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGGAGAGAGACCA	791
Db	734	CATACCCCCCAAGAAAAACAGCCATGGGATAAAAAACCTTACCTTCGGGGACCAAAACC	793
Qy	792	C 792	
Db	794	C 794	
RESULT 3			
BQ674807			
LOCUS			
DEFINITION BQ674807 668 bp mRNA linear EST 15-JUL-2002			
5', mRNA sequence.			
ACCESSION BQ674807			
VERSION BQ674807.1 GI:21785641			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 668)			

AUTHORS		NIH-MGC http://mgc.nci.nih.gov/.	
TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL		Unpublished (1999)	
COMMENT		Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LICM2377 row: c column: 09 High quality sequence stop: 656.	
FEATURES		Location/Qualifiers	
source		1..668 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6212672" /clone_lib="NIH_MGC_102" /tissue_type="epidermoid carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."	
BASE COUNT		216 a 148 c 167 g 137 t	
ORIGIN			
Query Match 51.2%; Score 525.4; DB 14; Length 668; Best Local Similarity 99.8%; Pred. No. 3e-114; Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	500	TGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTTGGCCCATCAG	559
Db	1	TGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTTGGCCCATCAG	60
Qy	560	GAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAACCCCGCAGGAGCAATGGCT	619
Db	61	GAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAACCCCGCAGGAGCAATAGCT	120
Qy	620	GCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAGAAA	679
Db	121	GCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAGAAA	180
Qy	680	TTTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATAAACA	739
Db	181	TTTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATAAACA	240
Qy	740	CTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGCCACTGGATCG	799
Db	241	CTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGCCACTGGATCG	300
Qy	800	GCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACT	859
Db	301	GCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACT	360
Qy	860	ACAAAATTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGCCAGGAGAG	919
Db	361	ACAAAATTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGCCAGGAGAG	420
Qy	920	ACTGTGCTGGGTTGATTATGCTGGGCAAGTGGACGATTTTCAATGTGAAGACGTCATA	979
Db	421	ACTGTGCTGGGTTGATTATGCTGGGCAAGTGGACGATTTTCAATGTGAAGACGTCATA	480
Qy	980	ACTTCATTTGGAAAAACAGAGGAGACAGTACTGTCTATCTGCATTA	1026
Db	481	ACTTCATTTGGAAAAACAGAGGAGACAGTACTGTCTATCTGCATTA	527

```
RESULT 4
AL568743/c
LOCUS      AL568743      967 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION AL568743 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE005YH04 3 prime
            , mRNA sequence.
ACCESSION  AL568743
VERSION    AL568743.1 GI:12923387
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 967)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             Location/Qualifiers
     source           1..967
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DE005YH04"
                     /clone_lib="LTI_FL002_PL1"
                     /lab_host="DH10B"
                     /note="organ: placenta; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector. Library was constructed by Life
                     Technologies. Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     filiang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      222 a      221 c      217 g      290 t      17 others
ORIGIN
Query Match      46.3%; Score 475.4; DB 9; Length 967;
Best Local Similarity 96.7%; Pred. No. 2.6e-10;
Matches 504; Conservative 3; Mismatches 9; Indels 5; Gaps 2;

QY 506 GGGTACAGGCATGCCAGGCCCCAGGGCCCCCGGCCCTCTCTGCCCATCAGGAGCGG 565
      |||
DB 898 GGGTACAGGCATGCCAGGCCCCAGGGCCCCCGGCCCTCTCTCTGCCCATCAGGAGCGG 843
      |||
QY 566 TGGTGGCCCTGGCCCTGCAGATGAGCCAAACCCCGGCACCGGAGGACAAATGGCTGCCCGC 625
      |||
DB 842 TTGTGCCCTTGGCCCTGCAGATGAGCCAAACCCCGGCACCGGAGGACAAATGGCTGCCCGC 783
      |||
QY 626 CTCACCTGGAAGAACTTCACAGACAATGCTACTATTTTCAGTTGAGAAGAAATTTTG 685
      |||
DB 782 CTCACCTGGAAGAACTTCACAGACAATGCTACTATTTTCAGTTGAGAAGAAATTTTG 723
      |||
QY 686 AGGATCAAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATATAAACACTAGAG 745
      |||
DB 722 AGGATCAAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATATAAACACTAGAG 663
      |||
QY 746 AGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGGCCACTGGATCGGCCTCA 805
      |||
DB 662 AGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGGCCACTGGATCGGCCTCA 603
      |||
QY 806 CAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACTACAAAA 865
      |||
DB 602 CAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACTACAAAA 543
      |||
QY 866 ATTGGAAGCTGGACAGCGGATAAAGTGGGGTTCATGGCCATGGCCAGGAGAGAGACTGTG 925
      |||
DB 542 ATTGGAAGCTGGACAGCGGATAAAGTGGGGTTCATGGCCATGGCCAGGAGAGAGACTGTG 484
      |||

QY 926 CTGGTTGATTTATCTGGGCAGTGGACGAGTTCCAATGTGAAGACGTCAATACTTCA 985
      |||
DB 483 CTGGTTGATTTATCTGGGCAGTGGACGAGTTCCAATGTGAAGACGTCAATACTTCA 424
      |||
QY 986 TTTCGAAAAACACAGAGGAGACAGTACTCTCATCTGCATTA 1026
      |||
DB 423 TTTCGAAAAACAGAGGAGACAGTACTCTCATCTGCATTA 383
      |||

RESULT 5
BI456109
LOCUS      BI456109      906 bp      mRNA      linear      EST 21-AUG-2001
DEFINITION BI456109 NCL_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5251888 5',
            mRNA sequence.
ACCESSION  BI456109
VERSION    BI456109.1 GI:15246765
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 906)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11636 row: b column: 17
            High quality sequence stop: 730.

FEATURES             Location/Qualifiers
     source           1..906
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5251888"
                     /clone_lib="NCL_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      247 a      228 c      256 g      175 t
ORIGIN
Query Match      45.1%; Score 462.6; DB 13; Length 906;
Best Local Similarity 81.7%; Pred. No. 2.8e-99;
Matches 597; Conservative 0; Mismatches 124; Indels 10; Gaps 5;

QY 1 ATGCAACAAGATTGATGAGTTCGAGGTTAGACACTGAAGTAGCCCAACTTATCAGTGAT 60
      |||
DB 123 ATGCAGCAACAGATGATGAGGTCAGAGTTAGACACTGAAGTAGCCCAACTTATCAGTGAT 182
      |||
QY 61 ATGGAAGAAATGAAGCTAGTAGCTCCAAAGCATGGTTCAGCTCATCAAGAAATTTTACAATA 120
      |||
DB 183 ATGGAAGAGATGAATGTTGACTTCCAAAGCAGCGTTCAGCTCATCAAGAACTTTACCAT 242
      |||
QY 121 CTACAGAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCGAGGACCCCT 180
      |||
DB 243 CTACAGAGTCCCTCCCGGGCCCCAGAGGTCCAAAGGTGACAGAGGATCTCAGGACCCCT 302
      |||
QY 181 GGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGAGGGGGGACCGCTGGACACCTGCCCT 240
      |||
DB 303 GTTCCAACTGGCAACAAGGGACAGAAAGGAGAGAGAGGGAGAGCGCTGGTCCACCTGSCCT 362
      |||
```

```

241 GCGGTGAGAGGCCCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGGCAAGGA 300
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GCGGTGAGAGGCGACAAATTGGACCACTGGCGCCCTCTCGAGAGCGTGGCGCAAGGA 422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  TCTAAGGCTCCAGGCGCCCAAGGCTCCCGTGGTTCCTCCCTGGAAGCCGCGCCCTCAG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  TCCAAAGGCTCACAGGGTCCCAAGGATCTCGTGGGTCCCGCAGTGAAGCCCTGCA 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  GCGCCAGTGGGAGCCCAAGGCGCCCGCCGACCAAGGAGGAGGACTCCCGGCGCT 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GGACTAGTGGGAGCCCAAGGAGCCAGGCTCCAGGCAAGGATGGACTCCCTGGCGCT 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  CAGGCGCTCTCGCTTCCAGGCACTTACAGGCACTGAGGCGCTGGAGCTACCTGGA 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  CAGGCGCTCTCGCTTCCAGGCACTTACAGGCACTGAGGCGCTGGAGCTACCTGGA 602
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  CCTGGGAGCTGCCAGGCTTCCCTGGGTACCAAGGATGCCAGGCGCCCGGCGCC 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  CCTGGGAGCTGCCAGGCTTCCCAAGGCTGCCAGGCGATGCCCTGGGCGCTAAGGCGCACTA 662
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  GCGCTCTCGGCGCCATCAGGAGCGGTGGTGGCGCTGCGCTGCGCAATGAGCCACCGC 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GG-CCTCAGGCGCTCAGGAGC-ATGAGGCGATTTGCTCGCAAAATGAACCAAGCCCA 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  GCACCGGAGGACAATGGC--TGCCCGCGCTCACTGGAAGAACTTCACA---GACAAATGCT 655
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GCATCAGAGGTTCAAGCGATGTCCTCTCACTGGAAGAACTTCACGGGATTAATGCT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  ACTATTTTCACTG--AAAGAAATTTTGAGGATGCAAGCTTTCTGTGAAGACA 712
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  ACTATTTTTCATGCAAAAGAAATTTTGAAGATTTGTAAGCGTTCTGTGAGAAA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  713 AGTCTTCACAT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  841 AATTTTCAAT 851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
LOCUS BE910803 580 bp mRNA linear EST 29-SEP-2000
DEFINITION 601661855F1 NCI_CGAP_Mam1 Mus musculus cdna clone IMAGE:3962292 5',
mRNA sequence.
ACCESSION BE910803
VERSION BE910803.1 GI:10407765
KEYWORDS EST, house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
TITILE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9128 row: m column: 13
High quality sequence stop: 580.
Location/Qualifiers
1. 580
/organism="Mus musculus"
/strain="FVB/N"
/db.xref="taxon:10090"
/clone="IMAGE:3962292"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"

```

```

/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 159 a 153 c 161 g 107 t
ORIGIN

Query Match 43.4%; Score 445.6; DB 12; Length 580;
Best Local Similarity 85.5%; Pred. No. 2.6e-95;
Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 290 CGCGCAAGAGTCTAAAGGCTCCAGGCGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGC 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GCAGCAAGAGTCCAAAGGCTCACAGGCTCCCAAGGATCTCGTGGTCCCCAGGGAAGC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 350 CGCGCTCAGGCGCCCGCAGTGGGAGCCAGGCGCCCGGCGCCAGGCAAGAGGAGAC 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CTGGCCCTCAAGGAGCTTAGTGGGAGCCAGGAGCCAGGATCCACAGGCAAGGATGGAC 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 410 TCCCGGCGCTCAGGCGCTCTCGCTTCCAGGGACTTCAGGGACCGTTGGGGAGCGCTG 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TCCCTGGCCCTCAGGCGCTCTCGCTTCCAGGGACTTCAGGGAGCTGTGGGTGGAGCGCTG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 470 GGGTGCCTGAGACCTCGGGGACTGCGAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCA 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GAGTACCTGGACCTCGGGGTTGCCAGGCTTGGCAGGGTGCCAGGCATGCTGGGCGCTA 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 530 AGGGCGCCCGCGCCCTCTCGGCGCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAAGATG 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AGGGACCACTTGGCCCTCCAGGCGCTCCAGGCGCTCAGGAGCAATGGAGCAATTTGCAAGAATG 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 590 AGCCAAACCCGCGAGGAGGAGATGCTGCCGCCCTCAGTGGAGAACTTCACAGACA 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AACCAACCCGAGCATCAGAGGTTCACGGATGTCCGCCCTCAGTGGAGAACTTCACAGATA 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 650 AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGAGCTTTTCTGTGAAG 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCTAAGCTTTTCTGTGAAG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 710 ACAAGCTTTCACATCTGTTTTCATAAACAACATAGAGAGGAGACAGCAATGGATAAAAAAC 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 ACAATCTTCCCATCTCGTTTTTCATAAACTCAAGAGAAAGACAGCAATGGATAAAAAAGC 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 770 AGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAAT 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 ATACCGTGGGAGAGAGAGGAGCCATTGGATCGGCTCACAGACTCAGAGCAAGGAAAGCGAAT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 830 GGAAGTGGCTGGATGGGAGATCTCCAGACTACAAAAATTG 869
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 GGAAGTGGCTAGAGGGGTCACTGTTGATTACAAAAACTG 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
LOCUS AW958053 552 bp mRNA linear EST 01-JUN-2000
DEFINITION EST370123 MAGE resequences, MAGE Homo sapiens cdna, mRNA sequence.
ACCESSION AW958053
VERSION AW958053.1 GI:8147736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITILE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cdna microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush

```

The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 115
 Seq primer: Reverse.

FEATURES

source
 1. .552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGE"
 /note="Vector: pBluescriptSKm"
 176 a 112 c 143 g 121 t

Query Match 42.6%; Score 437.2; DB 10; Length 552;
 Best Local Similarity 99.1%; Pred. No. 2.5e-93;
 Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 574 CTGGCCCTGCAGAAATGAGCAACCCCGGACCGGAGGACAATGGCTGCCCGCTCACTGG 633
 Db 1 CTGGCCCTGCAGAAATGAGCAACCCCGGACCGGAGGACAATGGCTGCCCGCTCACTGG 60
 QY 634 AAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCA 693
 Db 61 AAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCA 120
 QY 694 AAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATTAACACATAGAGAGGAACAG 753
 Db 121 AAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATTAACACATAGAGAGGAACAG 180
 QY 754 CAATGGATAAAAAACAAGATGTTAGGGAGAGAGCCACTGGATGGCCCTCAGACTCA 813
 Db 181 CAATGGATAAAAAACAAGATGTTAGGGAGAGAGCCACTGGATGGCCCTCAGACTCA 240
 QY 814 GAGCGTGAANAATGAATGGAATGGCTGATGGAGACATCTCCAGACTACAAAAATTTGGAAA 873
 Db 241 GAGCGTGAANAATGAATGGAATGGCTGATGGAGACATCTCCAGACTACAAAAATTTGGAAA 300
 QY 874 GCTGACAGCCGGATAACTGCGGTTCATGGCCATGGCCAGGAGAGACTGTCTGGGTG 933
 Db 301 GCTGACAGCCGGATAACTGCGGTTCATGGCCATGGCCAGGAGAGACTGTCTGGGTG 360
 QY 934 ATTTATGCTGGCAGTGGAAACGATTTTCCAATGTGAAGACGTCATTAAC-TTCATTTTGGCA 992
 Db 361 ATTTATGCTGGCAGTGGAAACGATTTTCCAATGTGAAGACGTCATTAAC-TTCATTTTGGCA 420
 QY 993 AAAACACAGGGAGACAGTACTGTCTCATCTGCATTA 1026
 Db 421 AAAACACAGGGAGACAGTACTGGCATCTGCATTA 454

RESULT 8
 BE290299
 LOCUS 601089246F1 NCI_CGAP_Mam5 Mus musculus cdna clone IMAGE:3484163 5',
 DEFINITION mRNA sequence.
 ACCESSION BE290299
 VERSION BE290299.1 GI:9171250
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 601)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cdna Library Preparation: Life Technologies, Inc.

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM8516 row: k column: 12
 High quality sequence stop: 551.
 Location/Qualifiers
 1. .601

FEATURES

source
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:3484163"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 156 a 161 c 179 g 105 t

ORIGIN

Query Match 40.5%; Score 415.8; DB 10; Length 601;
 Best Local Similarity 84.8%; Pred. No. 3.2e-88;
 Matches 501; Conservative 0; Mismatches 87; Indels 3; Gaps 3;
 QY 196 AAGGACACAGAAAGGAGAGAGGGGAGCCTGGACACACCTGGCCCTCGCGGTGAGAGAGGC 255
 Db 9 AGGTGACAGAAAGGAGAGAGGGGAGCCTGGTCCACCTGGCCCTCGCGGTGAGAGGGGC 68
 QY 256 CCAATTGGACACAGCTGTTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCACG 315
 Db 69 ACAATTGGACACAGCTGCGCCCTCTCTGGAGAGCGTGGCAGCAAGAGATCCAAAGGCTCACAG 128
 QY 316 GCGCCCAAGAGTCCCGTGGTTCCTCTGGG - AAGCCCGGCCCTCAGGGCCCGCCAGTGGGA 374
 Db 129 GGTCCCAAGAGTCTCGTGGTCCCGAGGTTAAGCTGGCCCTCAAGGACCTAGTGGGA 188
 QY 375 CCCAGGCCCCCGGCCACAGGCAAGAGAGGAGCTCCCGGCCCTCAGGGCCCTCTCTGG 434
 Db 189 CCCAGGACCAACAGTCCACAGGCAAGAGTGGCTCCCTGAGGCGCCCTCTCTGG 248
 QY 435 CTTCAGGAGACTTCAGGGCACCCTTGGGGAGCTGGGGTGCCTGGACCTCGGGGACTGCC 494
 Db 249 CTTCAGGAGACTTCAGGGCACCCTTGGGGAGCTGGGGTGCCTGGAGTACCTGGACCTCGGGGGTGGC 308
 QY 495 AGGCTTGGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGGGCC 554
 Db 309 AGGCTTGGC - AGGGTGCCAGGCATGCCCTGGGCTAAGGACCACCTGGCCCTCCAGGCC 367
 QY 555 ATCAGAGGGGTGGTGGCCCTCGAGAAATGAGCAACCCCGGCCAGCAGGAGACAA 614
 Db 368 CTCAGGAGCAATGGAGCCATTGGCTCTGCAGAAATGAACCAACCCAGCAGAGTCAA 427
 QY 615 TGGCTGCGCCCTCACTCGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAA 674
 Db 428 CGGATGCTCCCTCACTCGAAGAACTTCACAGTAAATGCTACTATTTTTCAGTTGAGAA 487
 QY 675 AGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTTTTCAT 734
 Db 488 AGAAA-TTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTTCCCATCTCTGTTTCAT 546
 QY 735 AAACACTAGAGAGAAACAGCAATGGATTAATAAACAGATGGTAGGAGAGA 785
 Db 547 AAACCTCAAGAGAGAAACAGCAATGGTAAATAAACATACGTGGGGAGAGA 597

RESULT 9
 B0955927
 LOCUS B0955927

957 bp mRNA linear EST 21-AUG-2002

DEFINITION AGENCOURT_8763247 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313505
5', mRNA sequence.
ACCESSION BQ955927
VERSION BQ955927.1 GI:22371405
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 957)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13740 Row: 1 Column: 13
High quality sequence stop: 665.

FEATURES
source
1..957
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6313500"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2-2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."

BASE COUNT	260 a	269 c	258 g	169 t	1 others
ORIGIN					
Query Match	39.3%	Score 403.4;	DB 14;	Length 957;	
Best Local Similarity	86.2%;	Pred. No. 3.3e-85;			
Matches 481;	Conservative	0;	Mismatches 72;	Indels 5;	Gaps 3;
Qy 1	ATGCAACAAGATTGTGTCAGGTCGAGGTAGACTGAAGTAGCCAACTTATCAGTGATT	60			
Dd 353	ATGCAGCAAGACATGATGAGGTCAAAGTTAGACTGAAGTGCCAACTTATCAGTGGTT	412			
Qy 61	ATGGAAGAATTAAGCTAGTAGTATCCAAGCATGGTCAGCTCATCAGAAATTTTACAATA	120			
Dd 413	ATGGAAGAGATGAACATCGTTGACTCCAAGCACGGTCAGCTCATCAAGAACTTTACCATT	472			
Qy 121	CTACAAGGTCCACCGGCCCCAGCGGTCCAGAGGTCACAGAGATCCCAGGGACCCCT	180			
Dd 473	CTACAAGGTCTCTCTGGCCCCAGAGGTTCCAAAAGGTGACAGAGGATCTCAGGGACCACT	532			
Qy 181	GGCCCAACTGCCAACAAAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCACTTGGCCCT	240			
Dd 533	GGTCCAACTGCCAACAGGGACAGAAAGGAGAGAGAGAGCGCTGGTCCACCTTGGCCCT	592			
Qy 241	CGGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGA	300			
Dd 593	CGGGGTGAGAGGGGCAAAATTGGACCAGTCGGCCCTCTCTGGAGAGCGTGGCAGCAAGGA	652			
Qy 301	TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG	360			
Dd 653	TCCAAAGGCTCACAGGTTCCCAAAGATCTCTGGGTCCCCAGGGAAGCTTGGCCCTCAA	712			
Qy 361	GGCCCAAGTGGGACCCAGGCCCCCGGCCGCCACAGGCAAGAGGACTCCCGGGCCT	420			
Dd 713	GGACTAGTGGGACCCAGGACCAAGGTCACCAAGCAGGATGACTCCCTTGGCCCT	772			
Qy 421	CAGGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGGCTTGGGGTGCCTGGA	480			

```
source
1..638
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730023E20"
/clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTAAATTAATCCCTCCCTCCG GAGAGAGAGATTCGAGTAAATTAATCCCTCCCTCCG 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATTAATCCCTCCCTCCG 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."
BASE COUNT      212 a 134 c 146 g 146 t
ORIGIN
Query Match      36.0%; Score 369.4; DB 10; Length 638;
Best Local Similarity 84.5%; Pred. No. 3.4e-77;
Matches 415; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 536 CCCCGGCCCTCTGCCCCCAGGAGCGGTGGTCCCTGCGCCCTGAGAAATGAGCCAA 595
Db 1 CACTTGGCCCTCCAGCCCTCAGAGCAATGAGCCATGGCTCTGAGAAATGAGCCAA 60
QY 596 CCCCGGCCCGGAGCAATGCTCCCGCCCTACTGGAAGAACTTCACAGCAAAATGCT 655
Db 61 CCCAGCATCAGAGTCAACGGATGTCGCCCTCAGTGAAGAACTTCACAGATAATGCT 120
QY 656 ACTATTTTTCAGTTGAGAAGAAATTTTGGAGATGCAAGCTTTCTGTGAAGACAAGT 715
Db 121 ACTATTTTTCAGTTGAGAAGAAATTTTGAAGATGCTAAGCTTTCTGTGAAGACAAT 180
QY 716 CTTCAATCTCTTTTCAACACTAGAGAGAAACAGCAATGGATGATAAAAAACAGATGG 775
Db 181 CTTCCACTCTCTTTTCAACACTAGAGAGAAACAGCAATGGATGATAAAAAACAGATCG 240
QY 776 TAGGAGAGAGAGCCACTGGATCGCCCTCAGACTCAGAGCGTGAATGAATGGAAT 835
Db 241 TGGGAGAGAGAAAGCCATTGGATCGCCCTCAGACTCAGAGCAAGAAAGCAATGGAAT 300
QY 836 GCTCGATGGGACATCTCCAGACTACAAAATTTGAAGCTGGACAGCCGGTAATCTGG 895
Db 301 GCTAGAGCGGTCACCTGTGTTATCAAAAAGCTGAAAAGCTGGACAAACAGATACTGG 360
QY 896 GTCAATGGCCATGGCCAGGAGAACTGTGCTGGGTGATTATGCTGGCAGTGGAAAG 955
Db 361 GCATGGCCATGGCCAGGAGAAAGCTGTGCTGGCTGATTATGCTGGCAGGACAGTGAAT 420
QY 956 ATTTCCAAATGTAAGACGCTCAATAACTTCAATTTGCGAAAGACAGGAGCAGTACTGT 1015
Db 421 ACTTCCAGTGTGATGAATCAATAACTTCAATTTGTGAGAAGGAAAGGAGGCGAGTACCAT 480
QY 1016 CATCTGCATTA 1026
Db 481 CATCCATATTA 491
RESULT 11
BM676508/c
LOCUS      500 bp      mRNA      linear      EST 27-FEB-2002
BM676508
```

```
DEFINITION      UI-E-EJ0-ahq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone
ACCESSION       BM676508
VERSION         BM676508.1 GI:18986404
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 500)
AUTHORS         Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE           Normalization and subtraction: two approaches to facilitate gene
JOURNAL         discovery
MEDLINE         Genome Res. 6 (9), 791-806 (1996)
COMMENT         97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seg primer: M13 Forward
POLYA-yes..
FEATURES         Location/Qualifiers
source          1..500
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="UI-E-EJ0-ahq-c-05-0-UI"
                /clone_lib="UI-E-EJ0"
                /tissue_type="fetal eyes, lens, eye anterior segment,
                optic nerve, retina, Retina Foveal and Macular, RPE and
                Choroid"
                /dev_stage="fetal and adult"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /note="Organ: eye; Vector: p7713-Pac (Pharmacia) with a
                modified polylinker; Site_1: EcoR I; Site_2: Not I;
                UI-E-EJ0 is a subtracted cDNA library constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed
                with an oligo-dT primer containing a Not I site. Double
                stranded cDNA was ligated to an EcoR I adaptor, digested
                with Not I, and cloned directionally into p7713-Pac
                vector. The oligonucleotide used to prime the synthesis of
                first-strand cDNA contains a library tag sequence that is
                located between the Not I site and the (dT)18 tail. The
                sequence tags for this library are: fetal eyes, AGAATCAAGA
                ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCGAT;
                optic nerve, CCATTAGTG; retina, CCGG; Retina Foveal and
                Macular, CTCG; RPE and Choroid, ACCGA. This library was
                created for the program, Gene Discovery in the Visual
                System, supported by National Eye Institute (NEI).
                TAG_LIB-UI-E-EJ0
                TAG_TISSUE-human eye anterior segment
                TAG_SEQ-AATGCCGCGAT"
BASE COUNT      109 a      89 g      179 t
ORIGIN
Query Match      34.2%; Score 351.4; DB 14; Length 500;
Best Local Similarity 99.7%; Pred. No. 5.9e-73;
Matches 352; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 674 AAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAAGTCTTCACATCTTGTTC 733
Db 494 AGGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAAGTCTTCACATCTTGTTC 435
QY 734 TAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGACC 793
```

```

|||||
Db 434 TAAACACTAGAGAGAACAGCAATGGATAAAAAACAAGATGTTAGGAGAGAGAGCCACT 375
QY 794 GGATGGCCTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTGGATGGAGCATCTC 853
Db 374 GGATGGCCTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTGGATGGAGCATCTC 315
QY 854 CAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAG 913
Db 314 CAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAG 255
QY 914 GAGAAGACTGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 973
Db 254 GAGAAGACTGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 195
QY 974 TCAATAAATTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 1026
Db 194 TCAATAAATTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 142

RESULT 12
LOCUS BM713891
DEFINITION UI-E-EJ0-ahq-c-05-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
ACCESSION BM713891
VERSION UI-E-EJ0-ahq-c-05-0-UI 5', mRNA sequence.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE discovery
COMMENT Genome Res. 6 (9), 791-806 (1996)
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msources@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 443-498, >POLY_A$Simple_repeat
Seq primer: M13 Reverse.
FEATURES
source
Location/Qualifiers
1..507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ahq-c-05-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac

```

```

vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT 184 a 89 c 121 g 112 t 1 others
ORIGIN
Query Match 34.1%; Score 350; DB 14; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.3e-72;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 677 AAATTTTTCAGCATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTTCATAA 736
Db 1 AAATTTTTCAGCATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTTCATAA 60
QY 737 ACACCTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCCACTGGA 796
Db 61 ACACCTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCCACTGGA 120
QY 797 TCGGCCTCACAGACTCAGAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 856
Db 121 TCGGCCTCACAGACTCAGAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 180
QY 857 ACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAGGAG 916
Db 181 ACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAGGAG 240
QY 917 AAGACTGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 976
Db 241 AAGACTGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
QY 977 ATAACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 1026
Db 301 ATAACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 350

RESULT 13
LOCUS BQ771366
DEFINITION UI-M-FI0-byu-g-09-0-UI.r1 NIH_BMAP_FI0 Mus musculus cDNA clone
ACCESSION BQ771366
VERSION BQ771366.1 GI:21979842
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE 1 (bases 1 to 808)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
source Location/Qualifiers
1..808
/organism="Mus musculus"
/strain="C57BL/6"

```

```
/db_xref="taxon:10090"
/clone_lib="IMAGE:5702432"
/clone_lib="NIH_BMAP_F10"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (11 phage resistant)"
Note="Organ: Brain; Vector: pX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCACACAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
```

BASE COUNT 270 a 158 c 173 g 206 t 1 others

Query Match 33.5%; Score 343.8; DB 14; Length 808;
Best Local Similarity 85.1%; Pred. No. 4.4e-71;
Matches 384; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 576 GGGCTGCGAGATGACGCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTGGAA 635
DB 1 GGCTCTGCGAGAATGAACCAACCCAGCATCAGAGTCAACGGATGTCGCCCTCACTGGAA 60

QY 636 GAACCTTCACAGCAAAATGCTACTATTTTCAGTTGGAAGAAGAAATTTTGGAGATGCCAA 695
DB 61 GAACTTCACAGTAATGCTACTATTTTCATTTGGAAGAAGAAATTTTGAAGATGCTAA 120

QY 696 GTTTTCTGTGAAGCAAGCTTCTACATCTGTTTTCATATAACACTAGAGAGAAACAGCA 755
DB 121 GCTTTTCTGTGAAGCAAAATCTCCCATCTGCTTTTTCATATAACTCAAGAGAAGACAGCA 180

QY 756 ATGGATAAAAAACAGATGGTAGGAGAGAGACCCACTGGATCGCGCTCACAGACTCAGA 815
DB 181 ATGGATAAAAAACATACCGTGGGAGAGAAAGCCATTGGATCGCGCTCACAGACTCAGA 240

QY 816 GCGTGAATAATGAATGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGGAAGC 875
DB 241 ACAGGAACCGAATGGAAGTGGCTAGACGGTCACTGTTGATTACAAAACTGGGAACG 300

QY 876 TGGACAGCCGGTAATCTGGGGTCAATGGCCATGGCCAGAGAGACTGTGCTGGTGTGAT 935
DB 301 TGGACAACCAAGATACTGGGGCAGTGGCCATGGCCAGAGAGACTGTGCTGGTGTGAT 360

QY 936 TTATGCTGGCAGTGAACGATTCCAAATGTGAACGTCATTAATCTTTCATTTGCGAAAA 995
DB 361 TTACGAGGACAGTGAATGACTTCCAGTGTGATGAATCAATAACTTCATTGTTGAGAA 420

QY 996 AGACAGGAGAGACGACTACTGTCATTCGCAATTA 1026
DB 421 GGGAAGGAGGAGGACGATACCATCATCATTA 451

RESULT 14
AA304251
LOCUS
DEFINITION
EST17006 Aorta endothelial cells, TNF alpha-treated Homo sapiens
CDNA 5' end similar to similar to Fc receptor for IgE, low affinity
, mRNA sequence.
ACCESSION
AA304251
VERSION
AA304251.1 GI:1956604
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 339)
Adams M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai C., Clayton, R.A.,
Cline, F.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon
, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and
Venter, J.C.

Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL
MEDLINE
COMMENT

Other ESTs: THC84598

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1. 339

source

/organism="Homo sapiens"

/db_xref="ATCC (inhost):115883"

/db_xref="taxon:9606"

/clone_lib="Aorta endothelial cells, TNF alpha-treated"

/cell_type="endothelial cell"

/dev_stage="adult"

/note="Organ: aorta; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI"

BASE COUNT 107 a 69 c 86 g 71 t 6 others

ORIGIN

Query Match 30.3%; Score 311; DB 9; Length 339;

Best Local Similarity 96.2%; Pred. No. 2.1e-63;

Matches 326; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

QY 573 CTTGCCCTGCGAGATGACCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTG 632

DB 1 CTTGCCCTGCGAGATGACCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTG 60

QY 633 GAAGAACTTCACAGCAAAATGCTACTATTTTCAGTTGAGAAGAAATTTTGGAGATGC 692

DB 61 GAAGAACTTCACAGCAAAATGCTACTATTTTCAGTTGAGAAGAAATTTTGGAGATGC 120

QY 693 AAAGCTTTTCTGTGAAGACAAGTCTTCACATCTGTTTTCATAAACAACACTAGAGAGAA 752

DB 121 AAAGCTTTTCTGTGAAGACAAGTCTTCACATCTGTTTTCATAAACAACACTAGAGAGAA 180

QY 753 GCAATGGATAAAAAACAGATGGTAGGAGAGAGACCCACTGGATCGGCTCACAGACTC 812

DB 181 GCAATGGATAAAAAACAGATGGTAGGAGAGAGACCCACTGGATCGGCTCACAGACTC 240

QY 813 AGAGGCTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAA 872

DB 241 AGAGGCTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAA 300

QY 873 ACCTGGACAGCCGGAT--AACTGGGGTCTATGCCATGGG 909

```

Qy      918  AGACTGTGCTGGGTTGATTGATTATGCTGG  944
      |||||
Db      316  AGACTGTGCTGGGTTGATTGATTATGCTGG  342

Search completed: March 21, 2003, 08:08:21
Job time : 1357.02 secs

```


GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 203.984 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_1695

Perfect score: 1026

Sequence: 1 atgcaacaagatttgatgag.....cagtactgtcatcgtcatta 1026

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1026	100.0	2024	21	AAA07697 Human collectin en
2	1026	100.0	2262	24	ABA97932 Human scavenger re
3	1026	100.0	2628	22	AAH43036 Nucleotide sequenc
4	1024.4	99.8	2005	24	ABL95574 Human angiogenesis
5	1024.4	99.8	2005	24	ABL88085 Human PR07223 cDNA
6	1024.4	99.8	2181	22	AAI60628 Human polynucleoti
7	1024.4	99.8	2318	22	AAI58842 Human polynucleoti
8	1024.4	99.8	2641	22	AAC56903 Human EXMAD-14 cod
9	1024.4	99.8	2929	24	ABQ92072 Human polynucleoti

10	1024.4	99.8	2930	20	AAV55746 Human secreted pro
11	883.8	86.1	1521	23	DNA encoding novel
12	802	78.2	2637	22	Nucleotide sequenc
13	414.8	40.4	2256	22	Nucleotide sequenc
14	169.4	16.5	493	23	DNA encoding novel
15	157.6	15.4	3685	20	Human CSR3 protein
16	157.6	15.4	3810	20	Human CSRI protein
17	150.2	14.6	873	23	DNA encoding novel
18	150.2	14.6	1062	23	DNA encoding novel
19	146.8	14.3	1329	23	DNA encoding novel
20	143	13.9	5041	22	Human nervous syst
21	121.6	11.9	756	14	Collagen-like poly
22	121.6	11.9	756	17	Collagen-like poly
23	121.4	11.8	5676	15	Human collagen (Ty
24	121.4	11.8	8284	22	Human breast cance
25	121.4	11.8	8284	22	Human breast cance
26	121.4	11.8	8284	22	Human breast cance
27	118.4	11.5	6109	23	Murine Col5a3 cDNA
28	118	11.5	756	14	Collagen-like poly
29	118	11.5	756	17	Collagen-like poly
30	117	11.4	9287	24	Human benign prost
31	115.8	11.3	4748	22	AD06573 Bovine alpha1(I) c
32	115	11.2	4270	24	Mouse ischaemic co
33	114.8	11.2	4816	24	Human cDNA differe
34	114.8	11.2	5058	23	DNA encoding novel
35	114.8	11.2	6358	22	Human EST-derived
36	114.8	11.2	6691	23	DNA encoding novel
37	114	11.1	5060	24	Prostate cancer-as
38	113.6	11.1	3181	19	Nucleotide sequenc
39	113.4	11.1	4821	23	DNA encoding novel
40	113.4	11.1	6200	23	Human COL5A3 cDNA
41	113.4	11.1	6368	24	Novel human coding
42	113.2	11.0	2319	13	DNA encoding human
43	113.2	11.0	2319	18	Human type IV coll
44	113.2	11.0	3898	22	Murine adipocytes-
45	112.4	11.0	4498	22	Porcine alpha2(I)

ALIGNMENTS

RESULT 1	
AAA07697	
ID	AAA07697 standard; DNA; 2024 BP.
AC	AAA07697;
XX	
DT	20-JUN-2000 (first entry)
XX	
DE	Human collectin encoding DNA.
XX	
KW	Collectin; human; antibacterial; antiviral; ds.
XX	
OS	Homo sapiens.
XX	
FT	Key Location/Qualifiers
FT	CDS 55..1698
FT	/*tag= a
FT	/product= "collectin"
FT	misc_feature 1..738
FT	/*tag= b
FT	/note= "fragment specifically claimed in claim 8"
FT	misc_feature 55..738
FT	/*tag= c
FT	/note= "fragment specifically claimed in claim 8"
FT	misc_feature 79..738
FT	/*tag= d
FT	/note= "fragment specifically claimed in claim 8"
FT	misc_feature 325..738
FT	/*tag= e
FT	/note= "fragment specifically claimed in claim 8"
FT	misc_feature 358..738
FT	/*tag= f

FT misc_feature /note= "fragment specifically claimed in claim 8"
FT 670..1695
FT /*tag= g
FT /note= "fragment specifically claimed in claim 5"
FT 685..738
FT /*tag= h
FT /note= "fragment specifically claimed in claim 7"
FT 730..738
FT /*tag= i
FT /note= "fragment specifically claimed in claim 7"
FT 739..1695
FT /*tag= j
FT /note= "fragment specifically claimed in claim 6"
FT 1696..2024
FT /*tag= k
FT /note= "fragment specifically claimed in claim 9"
XX
XX WO200011161-A1.
PN
XX
XX 02-MAR-2000.
PD
XX
XX 24-AUG-1999; 99WO-JP04552.
PF
XX
XX 24-AUG-1998; 98JP-0237611.
PR
XX
XX (FUSO) FUSO PHARM IND LTD.
PA
XX
XX Wakamiya N;
PI
XX
XX WPI: 2000-224596/19.
DR
XX P-PSDB; AAY77985.
DR
XX
XX New collectin of human origin having antibacterial and antiviral
PT activity, and gene encoding it useful for production of transgenic
PT animals and of antibodies for screening potential drug molecules -
PT
XX
XX Claim 5-9; Page 87-90; 106pp; Japanese.
PS
XX
XX The invention relates to polynucleotides encoding a new collectin of
CC human origin. The collectin can be used as an antibacterial and antiviral
CC agent and for screening potential drug molecules. The new collectin can
CC be produced by standard recombinant methodology. The present sequence
CC represents a DNA encoding the human collectin.
XX
XX Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;
SQ
Query Match 100.0%; Score 1026; DB 21; Length 2024;
Best Local Similarity 100.0%; Pred. No. 2.4e-251; Mismatches 0; Indels 0; Gaps 0;
Matches 1026; Conservative 0;
670 ATGCAACAAGATTGATGAGGTCGAGGTTAGACACTGAAAGTAGCCCAACTTATCAGTGATT 60
Db 670 ATGCAACAAGATTGATGAGGTCGAGGTTAGACACTGAAAGTAGCCCAACTTATCAGTGATT 729
QY 61 ATGGAGAATAAGCTAGTAGCTCAAGCATGTTGACAGTATCATCAAGAAATTTTACAATA 120
Db 730 ATGGAGAATAAGCTAGTAGCTCAAGCATGTTGACAGTATCATCAAGAAATTTTACAATA 789
QY 121 CTACAGGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACGAGGCCCT 180
Db 790 CTACAGGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACGAGGCCCT 849
QY 181 GGCCCAACTGGCAACAAGGAGCAGAAAGGAGAGGGGGAGCCCTGGACACCTGGCCCT 240
Db 850 GGCCCAACTGGCAACAAGGAGCAGAAAGGAGAGGGGGAGCCCTGGACACCTGGCCCT 909
QY 241 GCGGGTGAGAGAGGCCCAATTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGA 300
Db 910 GCGGGTGAGAGAGGCCCAATTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGA 969
QY 301 TCTAAGGCTCCACGGGGCCCCAAAGGCTCCCGTGGTTCCTGGGAAGCCCGGCTCAG 360
Db 970 TCTAAGGCTCCACGGGGCCCCAAAGGCTCCCGTGGTTCCTGGGAAGCCCGGCTCAG 1029

QY 361 GCCCCAGTGGGAGCCAGGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCCT 420
Db 1030 GCCCCAGTGGGAGCCAGGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCCT 1089
QY 421 CAGGGCCCTCCCTGGCTTCAGGGACTTCAGGGACCGTGGGGAGCCCTGGGCTGCTGA 480
Db 1090 CAGGGCCCTCCCTGGCTTCAGGGACTTCAGGGACCGTGGGGAGCCCTGGGCTGCTGA 1149
QY 481 CCTCGGGACTGCCAGGCTTGCTGGGGTACCAGGATGCCAGGCCGCCCAAGGCCGCCCT 540
Db 1150 CCTCGGGACTGCCAGGCTTGCTGGGGTACCAGGATGCCAGGCCGCCCAAGGCCGCCCT 1209
QY 541 GGCCTCTCTGGCCCATCAGGAGCGGTGTGCCCTGGCCCTGCAGAATGAGCAACGCC 600
Db 1210 GGCCTCTCTGGCCCATCAGGAGCGGTGTGCCCTGGCCCTGCAGAATGAGCAACGCC 1269
QY 601 GCACGGGAGGACAATGGCTGCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660
Db 1270 GCACGGGAGGACAATGGCTGCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1329
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGCAAGACAAGTCTTCA 720
Db 1330 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGCAAGACAAGTCTTCA 1389
QY 721 CATCTGTTTTCATAAACAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 780
Db 1390 CATCTGTTTTCATAAACAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 1449
QY 781 AGAGAGAGCCACTGGATGGCCCTCACAGACTCAGAGCGTGAAAATGAATGGAAGTGGCTG 840
Db 1450 AGAGAGAGCCACTGGATGGCCCTCACAGACTCAGAGCGTGAAAATGAATGGAAGTGGCTG 1509
QY 841 GATGGACATCTCCAGACTACAAAATTCGAAAGCTGGACAGCCGGATAACTGGGGTCAT 900
Db 1510 GATGGACATCTCCAGACTACAAAATTCGAAAGCTGGACAGCCGGATAACTGGGGTCAT 1569
QY 901 GCCCATGGCCAGGAGAAAGCTGTGCTGGGTTGATTTATGCTGGGAGTGGACGATTTTC 960
Db 1570 GCCCATGGCCAGGAGAAAGCTGTGCTGGGTTGATTTATGCTGGGAGTGGACGATTTTC 1629
QY 961 CAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAGACAGGAGACACTGTCATCT 1020
Db 1630 CAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAGACAGGAGACACTGTCATCT 1689
QY 1021 GCATTA 1026
Db 1690 GCATTA 1695
RESULT 2
ABA97932
ID ABA97932 standard; cDNA; 2262 BP.
XX ABA97932;
AC ABA97932;
XX
XX 25-APR-2002 (first entry)
XX
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
XX Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
KW rheumatoid arthritis; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 33..2262
FT /*tag= a
FT /product= "scavenger receptor-like protein"
XX JP2001340089-A.
XX
XX 11-DEC-2001.
PD

PA (BAKE/) BAKER K P.
 PA (FERR/) FERRARA N.
 PA (GERB/) GERBER H.
 PA (GERR/) GERRITSEN M E.
 PA (GODD/) GODDARD A.
 PA (GODO/) GODOWSKI P J.
 PA (GURN/) GURNEY A L.
 PA (HILL/) HILLAN K J.
 PA (MARS/) MARSTERS S A.
 PA (PANJ/) PAN J.
 PA (PAON/) PAONI N F.
 PA (STEP/) STEPHAN J F.
 PA (WATA/) WATANABE C K.
 PA (WILL/) WILLIAMS P M.
 PA (WOOD/) WOOD W I.
 XX
 PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 XX
 DR WPI: 2002-171999/22.
 DR P-PSDB: ABB95436.
 XX
 PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 PS Claim 1; Fig 27; 567pp; English.
 XX
 CC The present invention provides the protein and coding sequences of human
 CC PRO proteins. These are useful for treating or diagnosing a
 CC cardiovascular, endothelial or angiogenic disorder, including cardiac
 CC hypertrophy, trauma, cancer, age-related macular degeneration,
 CC atherosclerosis, hypertension, arterial stenosis, rheumatoid arthritis,
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 CC healing. The present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Query Match 99.8%; Score 1024.4; DB 24; Length 2005;
 Best Local Similarity 99.9%; Pred. No. 6.1e-251;
 Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAACAAGATTGATGAGGTTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
 DB 478 ATGCAACAAGATTGATGAGGTTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 537
 QY 61 ATGGAAGAAATGAAGTATGATCTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATA 120
 DB 538 ATGGAAGAAATGAAGTATGATCTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATA 597
 QY 121 CTACAAGTCCACCGGGCCCGAGGGTCCCAAGAGTGCAGAGGTCCAGGAGCCCGCCCT 180
 DB 598 CTACAAGTCCACCGGGCCCGAGGGTCCCAAGAGTGCAGAGGTCCAGGAGCCCGCCCT 657
 QY 181 GSCCAACTGGCAACAAGGGACAGAGAAGAGAGAGGGAGCCCTGGACCACTGGCCCT 240
 DB 658 GSCCAACTGGCAACAAGGGACAGAGAAGAGAGAGGGAGCCCTGGACCACTGGCCCT 717
 QY 241 GCGGTGAGAGAGGCCCAATTTGGACCACTGTGTCCTCCCGGAGAGCGTGGCGGCAAGA 300
 DB 718 GCGGTGAGAGAGGCCCAATTTGGACCACTGTGTCCTCCCGGAGAGCGTGGCGGCAAGA 777
 QY 301 TCTAAGGCTCCCGGGCCCGCAAGGCTCCCGTGTTCCTCGGGAAGCCCGGCCCTCAG 360
 DB 778 TCTAAGGCTCCCGGGCCCGCAAGGCTCCCGTGTTCCTCGGGAAGCCCGGCCCTCAG 837
 QY 361 GSCCCAGTGGGGACCCAGGCCCGCCCGCCAGGCAAGAGAGGACTCCCGGCCCT 420
 DB 838 GSCCCAGTGGGGACCCAGGCCCGCCCGCCAGGCAAGAGAGGACTCCCGGCCCT 897
 QY 421 CAGGGCCCTCTGCTTCCAGGGACTTCAGGGCACTTCAGGGCACTTCAGGGAGCCCTGGGTGCTTGA 480

DB 898 CAGGGCCCTCTGCTTCCAGGGACTTCAGGGCACTTCAGGGAGCCTGGGGTGGCTCTGA 957
 QY 481 CCTGGGAGACTGCCAGGCTTGCCTGGGTACCAGGCATGCCAGGCCCGCCAGGCCCGCC 540
 DB 958 CCTGGGAGACTGCCAGGCTTGCCTGGGTACCAGGCATGCCAGGCCCGCCAGGCCCGCC 1017
 QY 541 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG 600
 DB 1018 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG 1077
 QY 601 GCACGGAGGACAATGGCTGCCCGCTCAGTGAAGAATTCACAGACAAAATGCTACTAT 660
 DB 1078 GCACGGAGGACAATGAGTGGCCGCTCAGTGAAGAATTCACAGACAAAATGCTACTAT 1137
 QY 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA 720
 DB 1138 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA 1197
 QY 721 CATCTTTGTTTTCATAAACAACACAGAGGAACAGCAATGATAAAAAACAGATGGTAGG 780
 DB 1198 CATCTTTGTTTTCATAAACAACACAGAGGAACAGCAATGATAAAAAACAGATGGTAGG 1257
 QY 781 AGAGAGACCACTGGATCGGCTCAGACTCAGAGCTCAGAGCTGAAATGAAATGGAAGTGGCTG 840
 DB 1258 AGAGAGACCACTGGATCGGCTCAGACTCAGAGCTCAGAGCTGAAATGAAATGGAAGTGGCTG 1317
 QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGAATACCTGGGTCAT 900
 DB 1318 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGAATACCTGGGTCAT 1377
 QY 901 GGCATGGGCCAGGAGAAGACTGTCTGGTTGATTTATGCTGGCAGTGGGAACGATTTC 960
 DB 1378 GGCATGGGCCAGGAGAAGACTGTCTGGTTGATTTATGCTGGCAGTGGGAACGATTTC 1437
 QY 961 CAATGTGAAGCGTCAATTAACCTTCAATTTGCGAAAAGACAGGAGACAGTACTGTCATCT 1020
 DB 1438 CAATGTGAAGCGTCAATTAACCTTCAATTTGCGAAAAGACAGGAGACAGTACTGTCATCT 1497
 QY 1021 GCATTA 1026
 DB 1498 GCATTA 1503

RESULT 5
 ABL88085
 ID ABL88085 standard; cDNA; 2005 BP.
 XX
 AC ABL88085;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Human PR07223 cDNA sequence SEQ ID NO:27.
 XX
 KW Human; angiogenesis; cardiant; cytostatic; antilanglogenic; hypotensive;
 KW vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
 KW age-related macular degeneration; arterial stenosis; angina;
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
 KW wound healing; chromosome mapping; gene mapping; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200200690-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 20-JUN-2001; 2001WO-US19692.
 XX
 PR 23-JUN-2000; 2000US-213637P.
 PR 20-JUL-2000; 2000US-219556P.

XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 1045.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM39686.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1045; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Query Match 99.8%; Score 1024.4; DB 22; Length 2318;
Best Local Similarity 99.9%; Pred. No. 6.4e-251;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAACAGATTGATGAGTCGAGGTTAGACACTGAAGTAGCCCAACTTATCAGTGATT 60
DB 663 ATGCAACAGATTGATGAGTCGAGGTTAGACACTGAAGTAGCCCAACTTATCAGTGATT 722
QY 61 ATGGGAAGAATGAAGCTAGCTAGCTCCCAAGCATGTCAGCTCATCAAGAAATTTTCAATA 120
DB 723 ATGGGAAGAATGAAGCTAGCTAGCTCCCAAGCATGTCAGCTCATCAAGAAATTTTCAATA 782

QY 121 CTACAAGGTCCACCGGGCCCAAGGGGTCCAAGAGGTGACAGAGGATCCACGAGGACCCCT 180
DB 783 CTACAAGGTCCACCGGGCCCAAGGGGTCCAAGAGGTGACAGAGGATCCACGAGGACCCCT 842
QY 181 GCCCAACTGGCAACAAGGACAGAAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 240
DB 843 GCCCAACTGGCAACAAGGACAGAAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 902
QY 241 GGGGTGAGAGAGGCCCAATTTGGACACGTGCTGCCCGGAGAGCGTGGCGGCAAGA 300
DB 903 GGGGTGAGAGAGGCCCAATTTGGACACGTGCTGCCCGGAGAGCGTGGCGGCAAGA 962
QY 301 TCTAAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAG 360
DB 963 TCTAAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAG 1022
QY 361 GGGCCAGTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGGACTTCCCGGCCCT 420
DB 1023 GGGCCAGTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGGACTTCCCGGCCCT 1082
QY 421 CAGGGCCCTCTGGCTTCCAGGACTTCCAGGACCTTCAGGACCTTGGGGAGCCCTGGGTGCTGA 480
DB 1083 CAGGGCCCTCTGGCTTCCAGGACTTCCAGGACCTTGGGGAGCCCTGGGTGCTGA 1142
QY 481 CCTCGGGACTGCCAGGCTTGGCTGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCCC 540
DB 1143 CCTCGGGACTGCCAGGCTTGGCTGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCCC 1202
QY 541 GGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCACCCCG 600
DB 1203 GGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCACCCCG 1262
QY 601 GCACCGGAGGACAATGGCTGCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 660
DB 1263 GCACCGGAGGACAATGGCTGCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 1322
QY 661 TTTTCAGTTGAGAAAATAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTCA 720
DB 1323 TTTTCAGTTGAGAAAATAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTCA 1382
QY 721 CATCTGTGTTTCAATACACTAGAGGAGCAACAGCAATGATGATGATGATGATGATGATG 780
DB 1383 CATCTGTGTTTCAATACACTAGAGGAGCAACAGCAATGATGATGATGATGATGATGATG 1442
QY 781 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCTCAGAGCGTGAATAATGAATGGAAGTGTG 840
DB 1443 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGTG 1502
QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCA 900
DB 1503 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCA 1562
QY 901 GGCATGGCCAGGAGAGAGACTGCTGGGTGATTATGCTGGGACGTGGAACGATTC 960
DB 1563 GGCATGGCCAGGAGAGAGACTGCTGGGTGATTATGCTGGGACGTGGAACGATTC 1622
QY 961 CAATGTGAAGAGCTCAATAAATTTTCGAAAAAGACAGGAGAGAGAGAGAGAGAGAGAG 1020
DB 1623 CAATGTGAAGAGCTCAATAAATTTTCGAAAAAGACAGGAGAGAGAGAGAGAGAGAGAG 1682
QY 1021 GCATTA 1026
DB 1683 GCATTA 1688

RESULT 8
AAC66903
ID AAC66903 standard; cDNA; 2641 BP.
XX
AC AAC66903;
XX
DT 27-MAR-2001 (first entry)

XX Human EXMAD-14 coding sequence SEQ ID NO: 39.
DE
XX
KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
KW inflammation; reproductive disorder; cardiovascular disorder;
KW immune disorder; musculoskeletal disorder; developmental disorder;
KW gastrointestinal disorder; cell proliferation disorder; ss.
XX
OS Homo sapiens.
XX
PN WO20068380-A2.
XX
PD 16-NOV-2000.
XX
XX 10-MAY-2000; 2000WO-US12811.
XX
PF 11-MAY-1999; 99US-0133643.
XX
PR 23-AUG-1999; 99US-0150409.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
PI Azimzai Y;
XX
XX WPI: 2001-007395/01.
XX
DR P-PSDB; AAB27236.
XX
XX Isolated polynucleotide encoding extracellular matrix or
PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT preventing disorders associated with expression of EXMAD such as
PT proliferative, immune and genetic disorders -
XX
XX Claim 4; Page 121-122; 129pp; English.
XX
XX The present invention provides the protein and coding sequences for 25
CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC useful in the prevention and treatment of cancers, cell proliferation,
CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC gastrointestinal disorders and inflammation.
XX
XX Sequence 2641 BP: 823 A; 628 C; 608 G; 582 T; 0 other;
Query Match 99.8%; Score 1024.4; DB 22; Length 2641;
Best Local Similarity 99.9%; Pred. No. 6.8e-251;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAACAGATTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB
DB 912 ATGCAACAGATTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 971
QY 61 ATGGAAGAAATGAAGCTAGTAGACTTCCAAAGCATGCTCAGCTCATCAAGAATTTTACAATA 120
DB
DB 972 ATGGAAGAAATGAAGCTAGTAGACTTCCAAAGCATGCTCAGCTCATCAAGAATTTTACAATA 1031
QY 121 CTACAAGTCCACCGGCCCCCAGAGGTCCTCAAGAGTGACAGAGATCCCGAGGACCCCT 180
DB
DB 1032 CTACAAGTCCACCGGCCCCCAGAGGTCCTCAAGAGTGACAGAGATCCCGAGGACCCCT 1091
QY 181 GSCCAACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACCTGGCCCT 240
DB
DB 1092 GSCCAACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACCTGGCCCT 1151
QY 241 GCGGTGAGAGAGGCCCAATTTGACACAGCTGTCCTCCCGGAGAGCGTGGCGGCAAGGA 300
DB
DB 1152 GCGGTGAGAGAGGCCCAATTTGACACAGCTGTCCTCCCGGAGAGCGTGGCGGCAAGGA 1211
QY 301 TCTAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCGGAGCGCGGCCCTCAG 360
DB
DB 1212 TCTAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCGGAGCGCGGCCCTCAG 1271

QY 361 GGCCCCAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAGAGGAGACTCCCCGGCCCT 420
DB
DB 1272 GGCCCCAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAGAGGAGACTCCCCGGCCCT 1331
QY 421 CAGGCCCCCTCTGGCTTCAGGAGCTTCAGGACCCCTGGGGAGCCTGGGGTGGCTTGA 480
DB
DB 1332 CAGGCCCCCTCTGGCTTCAGGAGCTTCAGGACCCCTGGGGAGCCTGGGGTGGCTTGA 1391
QY 481 CCTCGGGGACTGCCAGGCTTGCTGGGTACCCAGGATGCCAGGCCCAAGAGGCCCCGCC 540
DB
DB 1392 CCTCGGGGACTGCCAGGCTTGCTGGGTACCCAGGATGCCAGGCCCAAGAGGCCCCGCC 1451
QY 541 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAATGAGCAACCCCG 600
DB
DB 1452 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAATGAGCAACCCCG 1511
QY 601 GCACCGGAGGACATGGCTGCCGCCCTCAGTGGAGAACTTCACAGACAAATGCTACTAT 560
DB
DB 1512 GCACCGGAGGACATGGCTGCCGCCCTCAGTGGAGAACTTCACAGACAAATGCTACTAT 1571
QY 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGCAAGTCTTCA 720
DB
DB 1572 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGCAAGTCTTCA 1631
QY 721 CATCTTGTTCATAAACAACACTAGAGAGGAACAGCAATGATGATGATGATGATGATG 780
DB
DB 1632 CATCTTGTTCATAAACAACACTAGAGAGGAACAGCAATGATGATGATGATGATGATG 1691
QY 781 AGAGAGCCACTGATGCGGCTCAGACTCAGAGCGTGAAGTGAAGTGAAGTGAAGTGA 840
DB
DB 1692 AGAGAGCCACTGATGCGGCTCAGACTCAGAGCGTGAAGTGAAGTGAAGTGAAGTGA 1751
QY 841 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGATGATGATGATGAT 900
DB
DB 1752 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGATGATGATGATGAT 1811
QY 901 GGCCTAGGCGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGCAACGATTT 960
DB
DB 1812 GGCCTAGGCGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGCAACGATTT 1871
QY 961 CAATGTGAAGCGTCAATACTTCAATTTGCGAAAAGACAGGAGACAGTACTGATCT 1020
DB
DB 1872 CAATGTGAAGCGTCAATACTTCAATTTGCGAAAAGACAGGAGACAGTACTGATCT 1931
QY 1021 GCATTA 1026
DB
DB 1932 GCATTA 1937
RESULT 9
ABQ92072
ID ABQ92072 standard; cDNA; 2929 BP.
XX
AC ABQ92072;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 69.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antiulcer; fungicide; antidiabetic; antihastmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
XX Homo sapiens.
XX

XX	23-MAR-1999	(first entry)	
DT	Human secreted protein clone bv227_1	coding sequence.	
DE			
XX			
XX	Secreted protein; human; nutritional supplements; immune stimulant;		
KW	immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;		
KW	activin/inhibin; chemokinesis; haemostasis; thrombolysis;		
KW	receptor/ligand activity; anti-inflammatory; tumour inhibitor;		
KW	cadherin/tumour invasion suppressor; ds.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	67..693	
FT		/*tag= a	
XX			
PN	W09855614-A2.		
XX			
PD	10-DEC-1998.		
XX			
PF	01-JUN-1998;	98WO-US11210.	
XX			
PR	29-MAY-1998;	98US-0087255.	
PR	04-JUN-1997;	97US-0868696.	
PR	04-JUN-1997;	97US-0868697.	
PR	04-JUN-1997;	97US-0868698.	
PR	04-JUN-1997;	97US-0868898.	
PR	04-JUN-1997;	97US-0868899.	
PR	04-JUN-1997;	97US-0868900.	
PR	04-JUN-1997;	97US-0869191.	
PR	04-JUN-1997;	97US-0869192.	
PR	04-JUN-1997;	97US-0869193.	
PR	04-JUN-1997;	97US-0869194.	
XX			
PA	(GEMV) GENETICS INST INC.		
XX			
PI	Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;		
PI	McCoy JM, Racie LA, Spaulding V, Treacy M;		
XX			
DR	WPI: 1999-059912/05.		
DR	P-FSDB: AAW73628.		
XX			
PT	New polynucleotides encoding secreted human proteins - derived from		
PT	human foetal brain, adult brain, foetal kidney, adult ovary, adult		
PT	retina, adult placenta or adult uterus cDNA libraries		
XX			
PS	Claim 26; Page 87-88; 127pp; English.		
XX			
CC	This sequence encodes a human secreted protein of the invention.		
CC	This DNA sequence was isolated from a human adult brain cDNA		
CC	library, and was designated clone bv227.1. The DNAs and proteins		
CC	are predicted to have biological activities which would make them		
CC	suitable for treating, preventing or ameliorating medical conditions in		
CC	humans and animals, although no supporting data is given. Suggested		
CC	activities include nutritional sources or supplements, immune		
CC	stimulating or suppressing activity, haematopoiesis regulating activity,		
CC	tissue growth activity, activin/inhibin activity,		
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,		
CC	receptor/ligand activity, anti-inflammatory activity, cadherin/tumour		
CC	invasion suppressor activity, and tumour inhibition activity. The DNAs		
CC	are also stated to be useful for gene therapy. A host cell transfected		
CC	with the DNA, or its subfragments and variants is useful for recombinant		
CC	production of the human secreted protein clones.		
XX			
SQ	Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;		
	Query Match 99.8%; Score 1024.4; DB 20; Length 2930;		
	Best Local Similarity 99.9%; Pred. No. 7e-251;		
	Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1 ATGCAACAAGATTGATGAGGTCGAGGTAGACACTGAAGTAGCCAACTTATCACTGATT 60		

DR P-PSDB; AAG63350.
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation
XX
XX Claim 4; Page 105-109; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
XX SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;
XX
XX Query Match 40.4%; Score 414.8; DB 22; Length 2256;
XX Best Local Similarity 87.0%; Pred. No. 1.3e-95;
XX Matches 468; Conservative 0; Mismatches 67; Indels 3; Gaps 1;
XX
XX QY 489 ACTGCCAGGCTTGCTGGGTACAGGATGCCAGGCGCCCAAGGCGCCCGGCCCTCC 548
XX Db 1393 ACTACAAGTCCACCGGCCCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGACCCCC 1452
XX QY 549 TGGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGGACCGGA 608
XX Db 1453 TGGCCCACTGTGCACAAGGGGACAGAAAGGAGAGAGGGGGGCGCTGGAC---CACCTGG 1509
XX QY 609 GGCAATGCTGCTCCGCTCACTGGAAGAACTTCACAGACAATGCTACTATTTTTCAGT 668
XX Db 1510 CCTCGGGCTCCGCTCACTGGAAGAACTTCACAGACAATGCTACTATTTTTCAGT 1569
XX QY 669 TGAGAAAGAAATTTTGGAGTGAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGT 728
XX Db 1570 TGAGAAAGAAATTTTGGAGTGAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGT 1629
XX QY 729 TTTTATAACACTAGAGAGGACAGCAATGGATATAAACAACAGATGTTAGGAGAGAG 788
XX Db 1630 TTTTATAACACTAGAGAGGACAGCAATGGATATAAACAACAGATGTTAGGAGAGAG 1689
XX QY 789 CCACCTGGATCGCCCTCACAGACTCAGAGCGTGAAGCTGAAATGAAGTGGCTGGGAC 848
XX Db 1690 CCACCTGGATCGCCCTCACAGACTCAGAGCGTGAAGCTGAAATGAAGTGGCTGGGAC 1749
XX QY 849 ATCTCAGACTACAAAATTTGAAAGCTGGACAGCGCGGATACCTGGGTCTATGGCCATGG 908
XX Db 1750 ATCTCAGACTACAAAATTTGAAAGCTGGACAGCGCGGATACCTGGGTCTATGGCCATGG 1809
XX QY 909 GCAGAGAGAACTGTGCTGGGTGATTTATCTGGGAGTGAACGATTTCCCAATGTGA 968
XX Db 1810 GCAGAGAGAACTGTGCTGGGTGATTTATCTGGGAGTGAACGATTTCCCAATGTGA 1869
XX QY 969 AGACGTCAATACTTCAATTTGGGAAAAAGACAGGAGACAGTCTCTCATCTGCATTA 1026
XX Db 1870 AGACGTCAATACTTCAATTTGGGAAAAAGACAGGAGACAGTCTCTCATCTGCATTA 1927
XX
XX RESULT 14
XX AAS71131/c
XX ID AAS71131 standard; cDNA; 493 BP.
XX
XX AC AAS71131;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #6935.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.
XX PN WO200175067-A2.
XX XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; AAG06944.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 1; SEQ ID No 6935; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 493 BP; 121 A; 116 C; 99 G; 157 T; 0 other;
XX
XX Query Match 16.5%; Score 169.4; DB 23; Length 493;
XX Best Local Similarity 97.7%; Pred. No. 2.4e-33;
XX Matches 214; Conservative 0; Mismatches 1; Indels 4; Gaps -4;
XX
XX QY 812 CAGACCGTGAATAATGAATGGAAGTGGCTGGATGGGACATC-TCCAGACTACA-AAAATG 869
XX Db 493 CAGACCGTGAATAATGAATGGAAGTGGCTGGATGGGACATC-TCCAGACTACA-AAAATG 434
XX QY 870 GAAAGCTGGAC-AGCGGATTAAT-AGGCTCATGGCCATGGCCAGGAGAGACATGTGCT 927
XX Db 433 GAAAGCTGGCCAGCCGGATTAAT-AGGCTCATGGCCATGGCCAGGAGAGACATGTGCT 374
XX QY 928 GGGTTGATTTATGCTGGCAGTGGAAACGATTTTCCAAATGTGAAGAGCTCAATCACTTCA 987
XX Db 373 GGGTTGATTTATGCTGGCAGTGGAAACGATTTTCCAAATGTGAAGAGCTCAATCACTTCA 314
XX QY 988 TCGGAAAAAGACAGGAGACAGTACTGTCTCATCTGCATTA 1026
XX Db 313 TCGGAAAAAGACAGGAGACAGTACTGTCTCATCTGCATTA 275
XX
XX RESULT 15
XX AAX27858

ID AAX27858 standard; DNA; 3685 BP.
XX AAX27858;
XX
XX
DT 02-JUN-1999 (first entry)
XX
DE Human CSR3 protein coding sequence.
XX
XX Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX
XX Homo sapiens.
XX
PN WO9909159-A1.
XX
PD 25-FEB-1999.
XX
XX 12-AUG-1998; 98WO-JP03602.
PF
XX 30-JUL-1998; 98JP-0230121.
PR
XX 13-AUG-1997; 97JP-0233396.
XX
PA (NLSB) JAPAN TOBACCO INC.
XX
XX Nakamura Y, Tokino T;
PI
XX WPI; 1999-181032/15.
DR P-PSDB; AAY00994.
XX
XX Scavenger receptor proteins - for treatment and diagnosis of
PT disorders involving cell stress
PT
XX
XX Claim 6; Page 142-150; 175pp; Japanese.
XX
CC This sequence encodes the human cellular stress response 3 (CSR3) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
XX Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other;
SQ

Query Match 15.4%; Score 157.6; DB 20; Length 3685;
Best Local Similarity 55.5%; Pred. No. 5.3e-30;
Matches 304; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

Qy 14 TGATGAGTCTGAGGTGACACTGAAGTAGCCAACTTATCATGATGATTATGAAGAATAATGA 73
Db 1406 TGCTCAGTGCCTGGCTGACCTCAAGCTCGGACCTCTCCATGATCGTGAGGAGATGA 1465

Qy 74 AGCTAGTAGACTCAAGCATGGTCAGTCATCAGAAATTTACATATCTACAGGTCCAC 133
Db 1466 AGGCAGTGGACACACAGCATGGAGAAATCTTCGCAATGTCCATCTCACCATCTCAGAGTGCCC 1525

Qy 134 CGGGCCCCAGGGTCCAGAGAGTGCACAGAGATCCACAGGACCCCTCGGCCCAACTGGCA 193
Db 1526 CCGGCCCTCCAGGACCAAGAGAGATTCAAGGAGATATGGGCGTGAAGGGCTGTGTGGCG 1585

Qy 194 ACAAGGGACAGAAAGGAGAGAGGGGAGCCTGGACCACTTGGCCCTCGGGTGTAGAGAG 253
Db 1586 GCAGAGGCCCGAAGGAGACCCCGCATCTTTGGGCCCTCGGACCCACAGGTCTCAGG 1645

Qy 254 GCCCAATTTGGACACAGCTGGTCCCCCGGAGAGCGTGGCGCAAGAGATCTAAAGGCTCCC 313
Db 1646 GGCAACCTGGAGAGCGCGGCTGTGGAGAAAGGGGCCCTGTGGCCCTCAGGGTTCC 1705

Qy 314 AGGGCCCCAAGAGTCCCGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGG 373
Db 1706 CAGGCCCTCAAGAGTCAAGAGGACAGCTTTGGAAGTGGAGGGCCGAGAGAGACGAGGCC 1765

Qy 374 ACCCAGGCCCCCGGGCCCCACCAGGCAAAAGAGGACTCCCGGGCCCTCAGGSCCCTCCTG 433
Db 1766 CAAAAGGGACATAGAGCCCGCCAGAGGCGCCCGGGGTCTCCAGGGCCCTCAG 1825

Qy 434 GCTTCCAGGGACTTTCAGGGACCGTTGGGAGACCTGGGGTSCCTGGACCTCGGGGACTGC 493
Db 1826 GGCTCAGGGAACCCGGGAATTGCAGGGAAGACAGAGGTTCACCAAGCCAGCGGGGGCCA 1885

Qy 494 CAGGCTTGCCTTGGGTACCAAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCCTGGGCC 553
Db 1886 TGGGGCCTAAGGTGAACCAAGGATCCAGGGTCCCTTGGTCTCCCGGGGCTCCAGGTC 1945

Qy 554 CATCAGGA 561
Db 1946 CACCAGGA 1953

Search completed: March 21, 2003, 02:19:27
Job time : 226.994 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:02:25 ; Search time 3431.41 Seconds
(without alignments)
11492.169 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_2024
Perfect score: 1355
Sequence: 1 atgcaacaagattgatgag.....agaaaaaggttatcccg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1339	98.8	2983	9	AB005145	AB005145 Homo sapi
2	1337.4	98.7	1886	9	HS0802985	AL713657 Homo sapi
3	1337.4	98.7	2005	6	AX454442	AX454442 Sequence
4	1337.4	98.7	2005	6	AX490920	AX490920 Sequence
5	1337.4	98.7	2641	6	AX047353	AX047353 Sequence
6	1337.4	98.7	3058	9	AB038518	AB038518 Homo sapi
7	895.2	66.1	2637	10	AB078434	AB078434 Homo sapi
8	887.4	65.5	3291	10	AB038519	AB038519 Mus muscu
9	722.4	53.3	4330	9	AB052103	AB052103 Homo sapi
c 10	488.8	36.1	169088	2	AC016128	AC016128 Homo sapi
c 11	488.8	36.1	188439	9	AP000915	AP000915 Homo sapi
c 12	484.6	35.8	71044	2	AC024368	AC024368 Homo sapi
c 13	466.2	34.4	187635	2	AP001022	AP001022 Homo sapi
c 14	465.8	34.4	178022	2	AP000900	AP000900 Homo sapi
c 15	437.4	32.3	188255	2	AP000939	AP000939 Homo sapi
c 16	388	28.6	182029	2	AC114677	AC114677 Mus muscu
c 17	388	28.6	193208	2	AC102618	AC102618 Mus muscu
c 18	378.4	27.9	130763	2	AC112416	AC112416 Rattus no
c 19	271	20.0	178022	2	AP000900	AP000900 Homo sapi
c 20	249	18.4	415	6	AX334493	AX334493 Sequence
c 21	164.4	12.1	187635	2	AP001022	AP001022 Homo sapi
c 22	157.6	11.6	3636	9	AB007829	AB007829 Homo sapi
c 23	157.6	11.6	3685	6	E32511	E32511 Scavenger r
c 24	157.6	11.6	3810	6	E32509	E32509 Scavenger r
c 25	150.6	11.1	2215	10	BC026446	BC026446 Mus muscu
c 26	129	9.5	6156	10	AB009993	AB009993 Mus muscu
c 27	129	9.5	188255	2	AP000939	AP000939 Homo sapi
c 28	128.2	9.5	2193	3	PALCOLIA	M25282 P.lividus a
c 29	125.8	9.3	6076	10	AF272661	AF272661 Rattus no
c 30	124.2	9.2	1041	4	AF138883	AF138883 Bos tauru
c 31	124.2	9.2	2053	5	AB008374	AB008374 Oncorhync
c 32	124.2	9.2	4502	5	AB052836	AB052836 Oncorhync
c 33	122.6	9.0	4995	10	MMU16789	U16789 Mus musculu
c 34	121.6	9.0	810	6	AR014116	AR014116 Sequence
c 35	121.4	9.0	4628	4	AB008683	AB008683 Bos tauru
c 36	121.4	9.0	5676	6	E07265	E07265 cDNA encodi
c 37	121.4	9.0	5676	9	HUMCALV	D90279 Human mRNa
c 38	121.4	9.0	7138	9	HUMPALV	M76729 Human pro-a
c 39	121	8.9	4581	4	ECU62528	U62528 Equus cabal
c 40	119.8	8.8	2010	9	HUMA2XICOL	L18987 Human alpha
c 41	119.4	8.8	5551	10	AF272662	AF272662 Rattus no
c 42	118.4	8.7	6109	6	AX239611	AX239611 Sequence
c 43	118.4	8.7	6109	10	AF176645	AF176645 Mus muscu
c 44	118	8.7	756	6	AR014110	AR014110 Sequence
c 45	118	8.7	6114	10	CRUPALV	M76730 Chinese ham

ALIGNMENTS

RESULT 1
LOCUS AB005145
DEFINITION Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1 GI:17026100
KEYWORDS Homo sapiens female tissue.lib:placenta cDNA to mRNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuh,A., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.

TITLE The membrane-type collectin CL-Pl is a scavenger receptor on vascular endothelial cells
JOURNAL J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE 21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani,K.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College, Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan [E-mail: ohtani@asahikawa-med.ac.jp, Tel: +81-166-68-2393, Fax: +81-166-68-2399]
FEATURES Location/Qualifiers

source 1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
gene 1..2983
/gene="CL-Pl"
CDS 71..2299
/gene="CL-Pl"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDDFAEEVEVQSFYKRFQIEQSTQCTKCKNNWALFKFSILLY ILCALLTITVAILGYKVVERMDNVTGGMETSRQYTDKLTAVESDLKLGIDQTKKAI STNSLSLFRQIDLDROOLREITEKTSKNKDTLEKQASGDALVDFTQSOLENN SFLITVNTKTLQYNGVYTNLQODTSVLQGNLQNMYSHNVVIMNLLNLTQVQORN LITNLSQSDTSQAIQRIKNDQNFQLOVFLQAKDDWLKQVQSLOTLANNSALA KANPTEEDMSQNSFQGMENITTSOANEQNLKQDLDLKHDAENRTAKFNQLEIE RFQFLETDVNIISISYTAHLRLTISLANEVNRTCTDITLTKHTDDTLISNNTLANI RLDVSRLMOQDLMRSRLDTEVANLSVIMEEMKLVDSKHGQLIKNFITLQPPGPRGP RGDGSGQPGPTNKGQKGEKGPDPGPPGAGERPIGPAGPPGERGKSGSGQPK GSRSPGKPGQSGDPGPPGKGLPGQPGPPGQGLQTVGPPGVPGRGLPG LPVGPMPKPGPPGPPGSCAVPLALQNEPTPAPENGGPPHKKNPTDKCYFSVE KEIREDAKLFEDKSSHLVFINTRREQQWIKQWVRESHWIGLTDSERENENKWLQD TSPDYKNMKAGQPNWGHGHPGEDCAGLIYAGQWNPQEDVNNFICEKDRITVLSS AL"

BASE COUNT	914 a	707 c	703 g	659 t
ORIGIN				
Query Match	98.8%; Score 1339; DB 9; Length 2983;			
Best Local Similarity	99.9%; Pred. No. 1.4e-300;			
Matches 1350; Conservative	0; Mismatches 0; Indels 1; Gaps 1;			
QY 1	ATGCAACAGATTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT	60		
DB 1271	ATGCAACAGATTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT	1330		
QY 61	ATGGAAGAAATGAAGCTAGTAGACTCCAGCATGCTGCTCATCAAGAATTTTACAATA	120		
DB 1331	ATGGAAGAAATGAAGCTAGTAGACTCCAGCATGCTGCTCATCAAGAATTTTACAATA	1390		
QY 121	CTACAGGTTCCACGGGCCCCAGGGTCCACAGAGGTGACAGAGGATCCACAGGACCCCT	180		
DB 1391	CTACAGGTTCCACGGGCCCCAGGGTCCACAGAGGTGACAGAGGATCCACAGGACCCCT	1450		
QY 181	GGCCCAACTGGCAACAGGCACAGAAAGAGAGAGGGGAGCCCTGGACCACTGGCCCT	240		
DB 1451	GGCCCAACTGGCAACAGGCACAGAAAGAGAGAGGGGAGCCCTGGACCACTGGCCCT	1510		
QY 241	GGGGTGTAGAGAGGCCCAATTGGACAGCTGGTCCCGCCGAGAGCGTGGCGGCAAGGA	300		
DB 1511	GGGGTGTAGAGAGGCCCAATTGGACAGCTGGTCCCGCCGAGAGCGTGGCGGCAAGGA	1570		
QY 301	TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAG	360		
DB 1571	TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAG	1630		
QY 361	GGCCCCAGTGGGAGCCAGGCCCCCGGCCACCACAGGCAAGAGGGACTCCCGGCCCT	420		
DB 1631	GGCCCCAGTGGGAGCCAGGCCCCCGGCCACCACAGGCAAGAGGGACTCCCGGCCCT	1690		

QY 421	CAGGCCCTCTCGGCTTCCAGGACTTCCAGGCACCGTTGGGAGCCCTGGGTGCTGGA	480		
DB 1691	CAGGCCCTCTCGGCTTCCAGGACTTCCAGGCACCGTTGGGAGCCCTGGGTGCTGGA	1750		
QY 481	CCTCGGAGCTGCCAGGCTTGCCTGGGTACCAGCATGCCAGGCCCCCAAGGCCCCCC	540		
DB 1751	CCTCGGAGCTGCCAGGCTTGCCTGGGTACCAGCATGCCAGGCCCCCAAGGCCCCCC	1810		
QY 541	GGCCCTCTCGGCTTCCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCCCAACCCCG	600		
DB 1811	GGCCCTCTCGGCTTCCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCCCAACCCCG	1870		
QY 601	GCACCGAGGACAATGGCTGCCCTCACTGGAAGAACTTCCACAGACAAATGCTACTAT	660		
DB 1871	GCACCGAGGACAATGGCTGCCCTCACTGGAAGAACTTCCACAGACAAATGCTACTAT	1930		
QY 661	TTTTTTCAGTTGAGAAAATAATTTTTGAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA	720		
DB 1931	TTTTTTCAGTTGAGAAAATAATTTTTGAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA	1990		
QY 721	CATCTTTTTCATAAACACTAGAGAGAAACAGCAATGATGATAAAAAACAGATGTTAGG	780		
DB 1991	CATCTTTTTCATAAACACTAGAGAGAAACAGCAATGATGATAAAAAACAGATGTTAGG	2050		
QY 781	AGAGAGGCCACTGGATCGGCTCACAGACTCAGAGCGGTGAAATGAAATGAAATGGAAGTGGCTG	840		
DB 2051	AGAGAGGCCACTGGATCGGCTCACAGACTCAGAGCGGTGAAATGAAATGAAATGGAAGTGGCTG	2110		
QY 841	GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGAATCACTGGGGTCAAT	900		
DB 2111	GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGAATCACTGGGGTCAAT	2170		
QY 901	GGCCATGGCCAGGAGAGACTGTCTGGTGTGATTTATGCTGGGAGTGAACGATTTTC	960		
DB 2171	GGCCATGGCCAGGAGAGACTGTCTGGTGTGATTTATGCTGGGAGTGAACGATTTTC	2230		
QY 961	CAATGTGAAGAGCTCAATAACTTTCATTTCCGAAAAGACAGGAGACAGTACTGTCATCT	1020		
DB 2231	CAATGTGAAGAGCTCAATAACTTTCATTTCCGAAAAGACAGGAGACAGTACTGTCATCT	2290		
QY 1021	GCATTTATACGGACTGTGATGGATCAGATGACAAATTTTTCAGCTCTCAAGGCAAGG	1080		
DB 2291	GCATTTATACGGACTGTGATGGATCAGATGACAAATTTTTCAGCTCTCAAGGCAAGG	2350		
QY 1081	ACACTCTCTTTCTAAATTCATCACCCTTCTCATCAGATTGAAAAAAGAAAGCACTGAAA	1140		
DB 2351	ACACTCTCTTTCTAAATTCATCACCCTTCTCATCAGATTGAAAAAAGAAAGCACTGAAA	2409		
QY 1141	CCAATTACTGAAAAAATTTGACAGTAGTGTGTTTTTACCATCCGTCATTTACCCAAAGAC	1200		
DB 2410	CCAATTACTGAAAAAATTTGACAGTAGTGTGTTTTTACCATCCGTCATTTACCCAAAGAC	2469		
QY 1201	TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTGTCACATGGACTGAA	1260		
DB 2470	TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTGTCACATGGACTGAA	2529		
QY 1261	TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTATACAAATTTATGCTTCCAAAGTA	1320		
DB 2530	TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTATACAAATTTATGCTTCCAAAGTA	2589		
QY 1321	TGGAACACTTCCCAATCAGAAAAAGTTATCAT	1351		
DB 2590	TGGAACACTTCCCAATCAGAAAAAGTTATCAT	2620		

RESULT 2
HSM802985
LOCUS Homo sapiens mRNA; cDNA DKFp547G1215 (from clone DKFp547G1215).
DEFINITION
ACCESSION AL713657
VERSION AL713657.1 GI:19584339
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1886)
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Weil, B. and Wiemann, S.
JOURNAL Direct Submission
TITLE Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFzp547G1215) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.
FEATURES
source
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFzp547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbri). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1886
/gene="DKFzp547G1215"
CDS
1..1256
/gene="DKFzp547G1215"
/note="similarity to asialoglycoprotein receptor (Mus musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKFNLEERFOLFETDIVNIISNTAHLRLTLNLEVRT TCTDILKTHDDTLTSLANTLANIRLDSVLRMOOILMRSLDTEVANLSVIMEKLV DSKHQILKFTILQGPGRGDRGSGQPGPTGNKGQKGEKGPBPAGRG PIGPAGPGRGKSGKSGPGKPGQSGSDPGPPGPGKGLPQGP PFGLOGTVEGPVGRGLPGLPGMPGPKPGPPGPPGAVPLALONPTPA PEDNGCPHWNFTDKCYFSEVEIFEAKLFCEKSKSHLVFNTRREEQWIKQWV GRESHWITGLTDSRENEKWLDTGTSPTDKNWKAGOPDNWGHGPGEDCAGLIYAGQW NDFOCEDVNNFICEKRETVLSSAL"
polyA_signal
1838..1843
/gene="DKFzp547G1215"
polyA_site
1869
/gene="DKFzp547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Query Match 98.7%; Score 1337.4; DB 9; Length 1886;
Best Local Similarity 99.9%; Pred. No. 3.2e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGCAACAAGATTTCATCAGGTTCAGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 228 ATGCAACAAGATTTCATCAGGTTCAGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 287
Qy 61 ATGGAAGAATGAAGCTAGTAGACTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATA 120
Db 288 ATGGAAGAATGAAGCTAGTAGACTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATA 347
Qy 121 CTACAAGTCCACCGGGCCCGCAGGGGTCCAAAGAGTGACAGAGATCCACAGGACCCCT 180
Db 348 CTACAAGTCCACCGGGCCCGCAGGGGTCCAAAGAGTGACAGAGATCCACAGGACCCCT 407
Qy 181 GGCCCACTGCCACAAAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 408 GGCCCACTGGCAACAGGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467

Qy 241 GCGGTGAGAGAGGCCCCAATTGGACAGCTGGTCCCGCGAGAGCGTGGCGGCAAGGA 300
Db 468 GCGGTGAGAGAGGCCCCAATTGGACAGCTGGTCCCGCGAGAGCGTGGCGGCAAGGA 527
Qy 301 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGCCCTCAG 360
Db 528 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGCCCTCAG 587
Qy 361 GGCCCACTGGGGAGCCAGGCCCCCGGCCCCACAGCAAGAGAGGACTCCCGGCCCT 420
Db 588 GCCTCAGTGGGGAGCCAGGCCCCCGGCCCCACAGCAAGAGAGGACTCCCGGCCCT 647
Qy 421 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCGCTGGGTGCTTGA 480
Db 648 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCGCTGGGTGCTTGA 707
Qy 481 CCTCGGGGACTGCCAGGCTTGCCTGGGTACAGGCATGCCAGGCCCAAGGGCCCGCCC 540
Db 708 CCTCGGGGACTGCCAGGCTTGCCTGGGTACAGGCATGCCAGGCCCAAGGGCCCGCCC 767
Qy 541 GGCCCTCTCTGGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCAACCCG 600
Db 768 GGCCCTCTCTGGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCAACCCG 827
Qy 601 GCACCGGAGGACAATGGCTGCCCGCTCCTCAGTGAAGAATTCACAGACAAATGCTACTAT 660
Db 828 GCACCGGAGGACAATGGCTGCCCGCTCCTCAGTGAAGAATTCACAGACAAATGCTACTAT 887
Qy 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGCAAGTCTTCA 720
Db 888 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGCAAGTCTTCA 947
Qy 721 CATCTTGTTCATAAACACATAGAGAGAACAGCAATGGATGATAAAACAGATGTTAGG 780
Db 948 CATCTTGTTCATAAACACATAGAGAGAACAGCAATGGATGATAAAACAGATGTTAGG 1007
Qy 781 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCTCAGAGCGTGAATAATCAATGGAAGTGGCTG 840
Db 1008 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCGTGAATAATCAATGGAAGTGGCTG 1067
Qy 841 GATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCCGGATTAATCGGGTCAAT 900
Db 1068 GATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCCGGATTAATCGGGTCAAT 1127
Qy 901 GSCCATGGCCAGGAGAGAGACTGTCTGGTGTGATTTATGCTGGGAGTGAACAGATTC 960
Db 1128 GSCCATGGCCAGGAGAGAGACTGTCTGGTGTGATTTATGCTGGGAGTGAACAGATTC 1187
Qy 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGAGGAGACAGTACTGTCAATCT 1020
Db 1188 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGAGGAGACAGTACTGTCAATCT 1247
Qy 1021 GCATTATACGAGCTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGGCAAGAG 1080
Db 1248 GCATTATACGAGCTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGGCAAGAG 1307
Qy 1081 ACACCTCTCTTCTAATTGATCACTCTCTCATCAGATTGAAAAAAGACACTGAAAA 1140
Db 1308 ACACCTCTCTTCTAATTGATCACTCTCTCATCAGATTG-AAAAAAGACACTGAAAA 1366
Qy 1141 CCAATTACTGAAAAAATTTGACAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
Db 1367 CCAATTACTGAAAAAATTTGACAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1426
Qy 1201 TTGGGAACATAATGTTCCCGAGGGTGATATGCTGATTTTTCATTTGTCACATGGACTGAA 1260
Db 1427 TTGGGAACATAATGTTCCCGAGGGTGATATGCTGATTTTTCATTTGTCACATGGACTGAA 1486
Qy 1261 TCACATAGATTCTCCTCGTTCAGTAACCGTGCAGTATATACAAATTTATGTCCTCCAAAGTA 1320
Db 1487 TCACATAGATTCTCCTCGTTCAGTAACCGTGCAGTATATACAAATTTATGTCCTCCAAAGTA 1546

QY 1321 TGAACACTCCAATCAGAAAAAGTTATCAT 1351
Db 1547 TGAACACTCCAATCAGAAAAAGTTATCAT 1577

RESULT 3
AX454442
LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 27 from Patent WO0208284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I.
and Ye,W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,
Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
I. (US)

FEATURES
source Location/Qualifiers
1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN

Query Match 98.7%; Score 1337.4; DB 6; Length 2005;
Best Local Similarity 99.9%; Pred. No. 3.3e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCACAAAGATTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 478 ATGCACAAAGATTGATAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 537
QY 61 ATGGGAAGAAATCAAGCTAGTAGCTCCAGCATGCTCAGCTCATCAAGAATTTTACAATA 120
Db 538 ATGGGAAGAAATCAAGCTAGTAGCTCCAGCATGCTCAGCTCATCAAGAATTTTACAATA 597
QY 121 CTACAAGGTCCACCGGCCCCAGGGTCCACAGAGGTGACAGAGGATCCCGAGGACCCCT 180
Db 598 CTACAAGGTCCACCGGCCCCAGGGTCCACAGAGGTGACAGAGGATCCCGAGGACCCCT 657
QY 181 GGCCTCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGACCCCTGGACCACTGGCCCT 240
Db 658 GGCCTCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGACCCCTGGACCACTGGCCCT 717
QY 241 GCGGGTGTAGAGAGGCCAATTTGACAGCTGCTCCCGCCGGAGAGCGTGGCGGCAAGGA 300
Db 718 GCGGGTGTAGAGAGGCCAATTTGACAGCTGCTCCCGCCGGAGAGCGTGGCGGCAAGGA 777
QY 301 TCTAAGAGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAG 360
Db 778 TCTAAGAGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAG 837
QY 361 GGCCTCACTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGAGGACTCCCGGCCCT 420
Db 838 GGCCTCACTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGAGGACTCCCGGCCCT 897
QY 421 CAGGCGCTCTCTGGCTTCCAGGAGCTTCAGGGCACCGTTGGGGAGCGTGGGGTGGCTGGA 480
Db 898 CAGGCGCTCTCTGGCTTCCAGGAGCTTCAGGGCACCGTTGGGGAGCGTGGGGTGGCTGGA 957

QY 481 CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGATGCCAGGCCCCCAAGGGCCCCCCC 540
Db 958 CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGATGCCAGGCCCCCAAGGGCCCCCCC 1017
QY 541 GGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAATCAGCAACCCG 600
Db 1018 GGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAATCAGCAACCCG 1077
QY 601 GCACCGGAGGACAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660
Db 1078 GCACCGGAGGACAATGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1137
QY 661 TTTTCAGTTGAGAAAAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGCAAGCTTCA 720
Db 1138 TTTTCAGTTGAGAAAAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGCAAGCTTCA 1197
QY 721 CATCTGTTTTCATAAACAACACTAGAGAGGAGCAACCAATGATAAAAAACAGATGTAGG 780
Db 1198 CATCTGTTTTCATAAACAACACTAGAGAGGAGCAACCAATGATAAAAAACAGATGTAGG 1257
QY 781 AGAGAGAGCCACTGGATCGGCCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 840
Db 1258 AGAGAGAGCCACTGGATCGGCCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 1317
QY 841 GATGGACATCTCCAGACTTACAAAATTTGAAAGCTGGACAGCCGGATTAACCTGGGGTCA 900
Db 1318 GATGGACATCTCCAGACTTACAAAATTTGAAAGCTGGACAGCCGGATTAACCTGGGGTCA 1377
QY 901 GGCCATGGCCAGGAGAGAGACTGCTGGGTTGATTTATGCTGGCAGTGGGAACGATTTC 960
Db 1378 GGCCATGGCCAGGAGAGAGACTGCTGGGTTGATTTATGCTGGCAGTGGGAACGATTTC 1437
QY 961 CAATGTGAAGAGCTCAATAAATTTTCGAAAAAGACAGGAGACAGTACTGTCTATCT 1020
Db 1438 CAATGTGAAGAGCTCAATAAATTTTCGAAAAAGACAGGAGACAGTACTGTCTATCT 1497
QY 1021 GCATTATAACGGAGCTGTGATGGATCAGATGAGCAAAATTTTCAGCTCTCAAAGGCAAGG 1080
Db 1498 GCATTATAACGGAGCTGTGATGGATCAGATGAGCAAAATTTTCAGCTCTCAAAGGCAAGG 1557
QY 1081 ACATCTCTTTCTAATTCATCACCCTTCATCAGATGAAAAAAGAAAGCACTGAAAA 1140
Db 1558 ACATCTCTTTCTAATTCATCACCCTTCATCAGATGAAAAAAGAAAGCACTGAAAA 1616
QY 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTCTTACCATCCGCTATTACCCAAAGAC 1200
Db 1617 CCAATTACTGAAAAAATTTGACAGCTAGTGTCTTACCATCCGCTATTACCCAAAGAC 1676
QY 1201 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 1260
Db 1677 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 1736
QY 1261 TCACATAGATTCCTCCGTCAGTAACCGTGGGATATACAAATTTATGTCCTTCCAAAGTA 1320
Db 1737 TCACATAGATTCCTCCGTCAGTAACCGTGGGATATACAAATTTATGTCCTTCCAAAGTA 1796
QY 1321 TGGAACTCTCCCAATCAGAAAAAGTTATCAT 1351
Db 1797 TGGAACTCTCCCAATCAGAAAAAGTTATCAT 1827

RESULT 4
AX490920
LOCUS AX490920 2005 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 27 from Patent WO0200690.
ACCESSION AX490920
VERSION AX490920.1 GI:22323797
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1
AUTHORS Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A., Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J., Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I. and Ye,W.
TITLE Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis
JOURNAL Patent: WO 0200690-A 27 03-JAN-2002;
Genentech, Inc. (US)
FEATURES Location/Qualifiers
Source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Query Match 98.7%; Score 1337.4; DB 6; Length 2005;
Best Local Similarity 99.9%; Pred. No. 3.3e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGCAACAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 478 ATGCAACAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 537
Qy 61 ATGGAAGAAATGAAGCTAGTAGCTCCAAAGCATGTGTAGCTCATCAAGAAATTTTACAATA 120
Db 538 ATGGAAGAAATGAAGCTAGTAGCTCCAAAGCATGTGTAGCTCATCAAGAAATTTTACAATA 597
Qy 121 CTACAAGTCCACGGGCCCCAGGGTCCAAAGGTGACAGAGTGACAGAGTATCCAGGGACCCCT 180
Db 598 CTACAAGTCCACGGGCCCCAGGGTCCAAAGGTGACAGAGTGACAGAGTATCCAGGGACCCCT 657
Qy 181 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGAGCCACTGGCCCT 240
Db 658 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGAGCCACTGGCCCT 717
Qy 241 GCGGTGAGAGAGGCCCAATTTGGACCAAGCTGTGTCCTCCCGGAGAGCGTGGGGGCAAGGA 300
Db 718 GCGGTGAGAGAGGCCCAATTTGGACCAAGCTGTGTCCTCCCGGAGAGCGTGGGGGCAAGGA 777
Qy 301 TCTAAGGCTCCAGGGCCCCAAGGCTCCGCTGTTCCCTGGCAAGCCCGGCCCTCAG 360
Db 778 TCTAAGGCTCCAGGGCCCCAAGGCTCCGCTGTTCCCTGGCAAGCCCGGCCCTCAG 837
Qy 361 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCACACAGCAAGAGGGACTCCCGGCCCT 420
Db 838 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCACACAGCAAGAGGGACTCCCGGCCCT 897
Qy 421 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCCTTGGGGAGCCCTGGGGTGCTGGA 480
Db 898 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCCTTGGGGAGCCCTGGGGTGCTGGA 957
Qy 481 CCTGGGAGCTGCCAGGCTTCGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 540
Db 958 CCTGGGAGCTGCCAGGCTTCGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 1017
Qy 541 GGCCCTCTCTGGCCATCAGGAGCGGTGTGTCCTGGCCCTGCAGAAATGAGCAACCCCG 600
Db 1018 GGCCCTCTCTGGCCATCAGGAGCGGTGTGTCCTGGCCCTGCAGAAATGAGCAACCCCG 1077
Qy 601 GCACGGAGGACAATGGCTGCCCGCTCAGTGGAAAGAACTTACAGACAAATGCTACTAT 660
Db 1078 GCACGGAGGACAATGAGCTGCCCGCTCAGTGGAAAGAACTTACAGACAAATGCTACTAT 1137
Qy 661 TTTTCAGTTGAAAGAAATTTTTCAGGATCAAGCTTTTCTGTGAAGACAAGTCTTCA 720
Db 1138 TTTTCAGTTGAAAGAAATTTTTCAGGATCAAGCTTTTCTGTGAAGACAAGTCTTCA 1197
Qy 721 CATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGTAGGG 780
Db 1198 CATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGTAGGG 1257
Qy 781 AGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 840

Db 1258 AGAGAGAGCCACTGATCGGCTCTCAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 1317
Qy 841 GATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACACCCGGATAACTTGGGGTCAT 900
Db 1318 GATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACACCCGGATAACTTGGGGTCAT 1377
Qy 901 GGCCATGGCCAGGAGAAAGACTGTGCTGGGTTGATTATGCTGGGAGTGGAAACGATTC 960
Db 1378 GGCCATGGCCAGGAGAAAGACTGTGCTGGGTTGATTATGCTGGGAGTGGAAACGATTC 1437
Qy 961 CAATGTGAAGAGCTCAATAAATTTTCGCGAAAAAGACAGGAGACAGTACTGTCACT 1020
Db 1438 CAATGTGAAGAGCTCAATAAATTTTCGCGAAAAAGACAGGAGACAGTACTGTCACT 1497
Qy 1021 GCATTATACGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAAAGGCAAGG 1080
Db 1498 GCATTATACGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAAAGGCAAGG 1557
Qy 1081 ACACCTCTTTCTAATTGATCACCCTTCTCATCAGATTGAAAAAAGAACACACTGAAAA 1140
Db 1558 ACACCTCTTTCTAATTGATCACCCTTCTCATCAGATTG-AAAAAAGAACACACTGAAAA 1616
Qy 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
Db 1617 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1676
Qy 1201 TTGGGAACATAAATCTCCCGAGGTGATGCTGATTTTCATTGTCACATGGACTGAA 1260
Db 1677 TTGGGAACATAAATCTCCCGAGGTGATGCTGATTTTCATTGTCACATGGACTGAA 1736
Qy 1261 TCACATAGATTCTCCTCGTCAGTAACCGTCGATTATACAAATTTATGTTCTTCCAAAGTA 1320
Db 1737 TCACATAGATTCTCCTCGTCAGTAACCGTCGATTATACAAATTTATGTTCTTCCAAAGTA 1796
Qy 1321 TGGACACTCCCAATCAGAAAAAGGTTATCAT 1351
Db 1797 TGGACACTCCCAATCAGAAAAAGGTTATCAT 1827
RESULT 5
AX047353 AX047353 2641 bp DNA linear PAT 15-DEC-2000
LOCUS Sequence 39 from Patent WO0068380.
DEFINITION AX047353
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 2641)
Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R., Lu,D.A. and Azimzai,Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
Incyte Genomics, Inc. (US)
FEATURES Location/Qualifiers
Source 1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CB1"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN
Query Match 98.7%; Score 1337.4; DB 6; Length 2641;
Best Local Similarity 99.9%; Pred. No. 3.3e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGCAACAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 912 ATGCAACAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 971

QY 61 ATGGAAGAAATGAAGCTAGTAGCTCCCAAGCATGTGTACGCTCATCAAGAATTTTCAATA 120
Db ATGGAAGAAATGAAGCTAGTAGCTCCCAAGCATGTGTACGCTCATCAAGAATTTTCAATA 1031
QY 121 CTACAGGTCCACCGGCCCCAGGGGTCCAAGAGGTGACAGAGGATCCCAGGACCCCT 180
Db CTACAGGTCCACCGGCCCCAGGGGTCCAAGAGGTGACAGAGGATCCCAGGACCCCT 1091
QY 181 GGCCCAACTGGCAACAAGGAGCAGAAAAGGAGAAAGGGGAGCCCTGGACACCTGGCCCT 240
Db GGCCCAACTGGCAACAAGGAGCAGAAAAGGAGAAAGGGGAGCCCTGGACACCTGGCCCT 1151
QY 241 GCGGTGAGAGAGGCCAATTTGGACCAAGCTGTCTCCCGGAGAGCGTGGCGGCAAGGA 300
Db GCGGTGAGAGAGGCCAATTTGGACCAAGCTGTCTCCCGGAGAGCGTGGCGGCAAGGA 1211
QY 301 TCTAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTCTGGGAAGCCCGGCCCTCAG 360
Db TCTAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTCTGGGAAGCCCGGCCCTCAG 1271
QY 361 GGCCCCAGTGGGACCCAGGCCCCCCGGGCCACCAGGCAAAAGAGGACTCCCGGCCCT 420
Db GGCCCCAGTGGGACCCAGGCCCCCCGGGCCACCAGGCAAAAGAGGACTCCCGGCCCT 1331
QY 421 CAGGGCCCTCTCGGCTTCAGGGAATTCAGGCAACGTTGGGGAGCCCTGGGTGCTTGA 480
Db CAGGGCCCTCTCGGCTTCAGGGAATTCAGGCAACGTTGGGGAGCCCTGGGTGCTTGA 1391
QY 481 CCTCGGGACTCCAGGCTTGCTGGGTACCAAGCATGCCAGGATGCCAGGCCCCAGGCCCCCCC 540
Db CCTCGGGACTCCAGGCTTGCTGGGTACCAAGCATGCCAGGATGCCAGGCCCCAGGCCCCCCC 1451
QY 541 GGCCCTCTCTGGCCCTCAGGAGCGTGTGCTGGCCCTGCAGAAATGAGCAACCCCG 600
Db GGCCCTCTCTGGCCCTCAGGAGCGTGTGCTGGCCCTGCAGAAATGAGCAACCCCG 1511
QY 601 GCACGGAGGACAATGGTGGCCGCTCAGTGAAGAACTTCACAGACAAATGCTACAT 660
Db GCACGGAGGACAATGGTGGCCGCTCAGTGAAGAACTTCACAGACAAATGCTACAT 1571
QY 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAGCTTCA 720
Db TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAGCTTCA 1631
QY 721 CATCTGTTTTCATAACACTAGAGGAACCAATGATGATAAAAAACAGATGGTAGG 780
Db CATCTGTTTTCATAACACTAGAGGAACCAATGATGATAAAAAACAGATGGTAGG 1691
QY 781 AGAGAGCCACTGGATGGCCCTCAGACTCAGAGCGTGAATGAATGGAATGGCTG 840
Db AGAGAGCCACTGGATGGCCCTCAGACTCAGAGCGTGAATGAATGGAATGGCTG 1751
QY 841 GATGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCGATAACTGGGTCTAT 900
Db GATGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCGATAACTGGGTCTAT 1811
QY 901 GGCCATGGGCCAGGAGACTGTCTGGTGTGATTATGCTGGGCAAGGAGCAATGTC 960
Db GGCCATGGGCCAGGAGACTGTCTGGTGTGATTATGCTGGGCAAGGAGCAATGTC 1871
QY 961 CAATGTGAAGAGCTCAATTAATTCATTTGCGAAAAGACAGAGGACAGTACTGTCT 1020
Db CAATGTGAAGAGCTCAATTAATTCATTTGCGAAAAGACAGAGGACAGTACTGTCT 1931
QY 1021 GCATTATACGACGTGTGATGGATCACATGACAAATTTTCAGCTCTCAAGAGCAAGG 1080
Db GCATTATACGACGTGTGATGGATCACATGACAAATTTTCAGCTCTCAAGAGCAAGG 1991
QY 1081 ACACCTCTTTCTAATTTGATCATCACTTCTCATCAGATTTGAAAAAAGACACTGAAA 1140
Db ACACCTCTTTCTAATTTGATCATCACTTCTCATCAGATTTGAAAAAAGACACTGAAA 2050
QY 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGCTCATTACCCAAAGAC 1200

Db 2051 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGCTCATTACCCAAAGAC 2110
QY 1201 TTGGGAACATAAATGTTCCCGAGGTGATATGCTGATTTTTCATTGTCACATGGACTGAA 1260
Db TTGGGAACATAAATGTTCCCGAGGTGATATGCTGATTTTTCATTGTCACATGGACTGAA 2170
QY 1261 TCACATAGATTCTCTCCGTCAGTAACCGTCGATTATACAAATATGTTCTTCCAAAGTA 1320
Db TCACATAGATTCTCTCCGTCAGTAACCGTCGATTATACAAATATGTTCTTCCAAAGTA 2230
QY 1321 TGGACACTCCAATCAGAAAAAGGTTATCAT 1351
Db TGGACACTCCAATCAGAAAAAGGTTATCAT 2261
RESULT 6
AB038518 3058 bp mRNA linear PRI 08-MAR-2001
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type I, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE Homo sapiens tissue_lib:Placenta cdna to mRNA.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F. and Nakamura, T.
TITLE Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
AUTHORS Nakamura, K. and Nakamura, T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamura@onbich.med.osaka-u.ac.jp,
Tel: 81-6-6879-3783(ex.3783), Fax: 81-6-6879-3789)
FEATURES
source Location/Qualifiers
1. 3058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Placenta"
1. 3058
/gene="SRCL"
209. 2437
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type I"
/protein_id="BAB39147.1"
/db_xref="gi:13365515"
/translation="MKDDFAEEVEVQSFYKRFIOEGTCTCKNNWALFSSILLY
ILCALLTITVAILGVKVKMDNVNVTGDMETSRQYDDKLTAVEDSKLKLQDTGKAI
STNSLSTFRSDILDRLQOLREITEKTSKNKDTLEKLOASGDALVDROSLKETLENN
SFLITVNTKLTQANGVYTNLQDTSVLQNLQNMYSNVVIMNINLNLITVQORN
LITNLQSVDDTSQAQIRKNDQNFQLOAKKDDTLWKEKVSQDGLTAAANSALA
KANNDLEDMSNLSNFTGOMENTTISQANEQNLKDLQDLHKDAENRTAKFNQLEE
RFQLFETDINIIISNYSYTAHLRLTATLSQANEVTTCTDTLTKHTDDTLNNLTANI
RDSVSLRMOODLMKSLRDTVANLSVIMEEMKLVDSKHGOLIKNFTLQCGPGPRGP
RGDRSGPGPTGKQKGEKPGEPGAGERPIGPGAPGGERGKSGSGQPK
GSRSPKPGQGGSGDPGPGKEGLPGQPGPPGPGQGLQCTVGVGPGVPGRLPG
LPVPGMPGPGPGPGPSGAVFLAQNEPTTAPEDNSCPPHWNKNTDKICYFSVE
KEIFEDAKLFCEDKSSHLVFINTEEQWIKKQMWGRESHWIGLTDSSRENEWKLDG
TSPOYKNWKAQPDNWGHGPGEDCAGLIYAGQWDFQEDVNNF ICEKORETVLSS
AL"
polyA_signal 3020..3025
/gene="SRCL"
BASE COUNT 911 a 764 g 639 t

ORIGIN

Query Match 98.7%; Score 1337.4; DB 9; Length 3058;
Best Local Similarity 99.9%; Pred. No. 3.4e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 ATGCAACAAGATTGATGAGTGCAGGTTAGACACTGAAGTAGCCAACTTATCATGATT 60
Db 1409 ATGCAACAAGATTGATGAGTGCAGGTTAGACACTGAAGTAGCCAACTTATCATGATT 1468

Qy 61 ATGGAAGAATGAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAATTTTACATA 120
Db 1469 ATGGAAGAATGAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAATTTTACATA 1528

Qy 121 CTACAAGTCCACGGGCCCCAGGGGTCCAAGAGGTGACAGAGTCCACAGGAGCCGCCCT 180
Db 1529 CTACAAGTCCACGGGCCCCAGGGGTCCAAGAGGTGACAGAGTCCACAGGAGCCGCCCT 1588

Qy 181 GGCCCACTGGCAACAAGGGACAGAAAGGAGAGAAAGGGGAGCCTGGACACCTGGGCCCT 240
Db 1589 GGCCCACTGGCAACAAGGGACAGAAAGGAGAGAAAGGGGAGCCTGGACACCTGGGCCCT 1648

Qy 241 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCCCGGAGAGCGTGGGGCAAGGA 300
Db 1649 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCCCGGAGAGCGTGGGGCAAGGA 1708

Qy 301 TCTAAAGGCTCCACAGGGCCCCAAGGGCTCCCGTGGTCCCGTGGGAAGCGCGGCCCTCAG 360
Db 1709 TCTAAAGGCTCCACAGGGCCCCAAGGGCTCCCGTGGTCCCGTGGGAAGCGCGGCCCTCAG 1768

Qy 361 GGCCCCAGTGGGGACCCAGGCCGCCCGGGGCCACAGCAAGAGGAGACTCCCGGGGCCCT 420
Db 1769 GGCCCCAGTGGGGACCCAGGCCGCCCGGGGCCACAGCAAGAGGAGACTCCCGGGGCCCT 1828

Qy 421 CAGGSCCTCTGCTGCTCCAGGGACTTTCAGGSCACCGTTGGGAGCGCTGGGTCCTGGA 480
Db 1829 CAGGSCCTCTGCTGCTCCAGGGACTTTCAGGSCACCGTTGGGAGCGCTGGGTCCTGGA 1888

Qy 481 CCTGGGGACTGCCAGGCTTGCCTGGGTACAGGCACTGCGAGGCCCAAGGGCCCCCCC 540
Db 1889 CCTGGGGACTGCCAGGCTTGCCTGGGTACAGGCACTGCGAGGCCCAAGGGCCCCCCC 1948

Qy 541 GGCCCTCTGCGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCGAGATGAGCCAAACCCG 600
Db 1949 GGCCCTCTGCGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCGAGATGAGCCAAACCCG 2008

Qy 601 GCACGGAGGACATGGTGCCTGCCCTCCTGAGGACTTTCAGAGAACTGCTACTAT 660
Db 2009 GCACGGAGGACATGGTGCCTGCCCTCCTGAGGACTTTCAGAGAACTGCTACTAT 2068

Qy 661 TTTTCAGTTGAGAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCA 720
Db 2069 TTTTCAGTTGAGAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCA 2128

Qy 721 CATCTTGTGTTTATAAACAATAGAGGAACAGCAATGGATGATAAAAAACAGATGGTAGG 780
Db 2129 CATCTTGTGTTTATAAACAATAGAGGAACAGCAATGGATGATAAAAAACAGATGGTAGG 2188

Qy 781 AGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTG 840
Db 2189 AGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTG 2248

Qy 841 GATGGACATCTCCAGACTCAAAAAATTTGAAAGCTGACAGCGCGGATAACTGGGGTCA 900
Db 2249 GATGGACATCTCCAGACTCAAAAAATTTGAAAGCTGACAGCGCGGATAACTGGGGTCA 2308

Qy 901 GGCCATGGGCCAGGAGAGACTGCTGGGTGATTTATGCTGGGCGAGTGGAAAGATTC 960
Db 2309 GGCCATGGGCCAGGAGAGACTGCTGGGTGATTTATGCTGGGCGAGTGGAAAGATTC 2368

Qy 961 CAATGTGAAGACGCTCAATACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCACT 1020
Db 2369 CAATGTGAAGACGCTCAATACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCACT 2428

Qy 1021 GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAGGCAAGG 1080
Db 2429 GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAGGCAAGG 2488

Qy 1081 ACATCTCCTTTCTAATTGATGATCACCTTCTCATCAGATTGAAAAAAGGACTGAAAA 1140
Db 2489 ACATCTCCTTTCTAATTGATGATCACCTTCTCATCAGATTG-AAAAAAGGACTGAAAA 2547

Qy 1141 CCAATTACTGAAAAAATTCACAGCTAGTGTCTTTTACCATCGCTCATTACCCAAAGAC 1200
Db 2548 CCAATTACTGAAAAAATTCACAGCTAGTGTCTTTTACCATCGCTCATTACCCAAAGAC 2607

Qy 1201 TTGGGAACATAAATGTTCCCGCAGGGTGATATCTGATTTTCATTTGTCACACTGGACTGAA 1260
Db 2608 TTGGGAACATAAATGTTCCCGCAGGGTGATATCTGATTTTCATTTGTCACACTGGACTGAA 2667

Qy 1261 TCACATAGATTCTCCTCGGTGAGTAACCGTGGGATATACAAATATATCTTCCAAAGTA 1320
Db 2668 TCACATAGATTCTCCTCGGTGAGTAACCGTGGGATATACAAATATATCTTCCAAAGTA 2727

Qy 1321 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1351
Db 2728 TGGAACTCCCAATCAGAAAAAGGTTATCAT 2758

RESULT 7
AB078434
LOCUS AB078434 2637 bp mRNA linear ROD 18-JUL-2002
DEFINITION Mus musculus CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB078434
VERSION AB078434.1 GI:21901968
KEYWORDS
SOURCE Mus musculus tissue_lib:Liver cdna to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuoh,A., Sakamoto,T., Itabe,H., Suzutani,T., Ogasawara,M., Yoshida,I. and Wakamiya,N.
TITLE cDNA cloning of mouse CL-P1 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2637)
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuoh,A., Sakamoto,T. and Wakamiya,N.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College, Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp, Tel:81-166-68-2393, Fax:81-166-68-2399)

FEATURES
source
1..2637
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_lib="Liver"
1..2637
/gene="CL-P1"
92..2320
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAC05523.1"
/db_xref="GI:21901968"
/translation="MKDFAEEVEVQSGYKRFQIQGTQCTCKKNWALKFSLVLL
ILCALITVAILGKVVERKNDVNTDGMETSHQYTDNKLTAVESDLKKLGGQAKKAL
STNSLSTFRSLDLRLQQLQEIETKSKNKDTLEKQANGSLDVRDSQSLKETLQNN
SFLITVTKTLOAYNGVYVTLQDTSVQNGLSQSVVIMNLNLNTVQQRN
LISNLQGVDDTSIAIORIKNDONLOQVFOAKKDDDLKKEVQSLOTIAANNSALA
KANNTILEDMSQLSSETGQMDNTTISQNEQSLKLDLQDKDTEKNTAVKPSQLEE
RFQVFTDIVNIISNTSYTAHLRTLSNLDVNTCTDTRHTDLDTSNLTNLVNI
RLDISILSRMQDDMRKLTDEVANLSVYMEEMKLVDKSGQLTKNFTILQGPFGPRP
KDRGSQGPFGPTGNKGQKGEKGPFGPAGERGTIGVPPGERSGSGSGQSPK

SFLLTVNKTLOAYGYVNTLQODTNVLQNLQSQMYSQSVVINNLNLNLTOQVBR
 LLSNLQSDNLSLQTKQNDQNTQVQLQAKKDLDLWKQVQSOQNTVQAANLSALA
 KANNLTDNDSLSQTKQNDQNTISQANQSLKDLDLQKTDNRTVAQSLEEE
 RFOVETIVNLSISYTAHLRLTSLNLMVMTDTLRTDLTDLSUNNLVNI
 RLOSILRWQDMRSLKSTEVANLSYEMZMLKVDSHGLIKNFITLQCPGPRGP
 KGRSGQSPGPTGNDQKQGEKGPPEKGPAGEKPTVPQGPGRSGSKSGQSGPK
 GSRSGKQPGPOGSGDQKQGGKDLQPGQPGQFQGVGVSGVSGVSGRGLPG
 LPVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
 KELLQDAKLFQEDKSHSLVFINSRREQQVKKHVTGRESHWIGITDSEQSEKWLKD
 SPVDYNKWKAGQPDNWSGCHGPGEDCAGLIVAGQWNPQCDIENNFTEKEREAVPSS
 II.

[illegible]

CDS	3...1871		/gene="SRCL"		/codon_start=1		/product="scavenger receptor with C-type lectin type II"		/protein_id="BAB39148.1"		/db_xref="GI:1336553"									
			/translation="MKDDFAEEVEVSPFKRFGIHEGTCQPKCHNNWALKFSIILLY		ILCALLTTVAIIGYKVVKMDNTVGGMETSRQTVDDKLTAVESDLKLGQDTGKKAI		STNSLSFTFQIDLDROQLREITKTSKNDLTLEKLOASGDALVDROSOLKEKLENN		SFLITVNTKLTQAYNGYVTLNQDTSVLQGNLQNGMYSHNVYIMNINLNLTQVQORN		LITNLSRYDDTSQAIOIRIKDNFQNLQVFLQAKKDIDLKQVSLQTLAANSALA									
			RANNDLDMNSQLNSFTQHMENITTSQANEQNLKDLQDLKDAENRTAKFNQLEE		RPQFETDIVIINISITTAHLRLTSLNNEVTRCTDLTKHTDDLSLNNLTANI		RDLDSVLRMDQNTKDETVANLSMEEMKLVDKSHGHLKFNFTILQGGPGRGP		RGDRSGQPPGPTKNGQKGEKGGPPGAGRGPTGAGPPGREGRGKGSQSPK		GSRSPGPKPGPGSPGPPGPGPKGELPGPOGPGFQGLQGTGEGVPGPGRGLPG									
			LPGVPGMPGPKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG		LPGVPGMPGPKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG		LPGVPGMPGPKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG		LPGVPGMPGPKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG		LPGVPGMPGPKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG									
BASE COUNT	1186 a		1065 c		1035 g		1044 t													
ORIGIN																				
Query Match	53.3%;		Score 722.4;		DB 9;		Length 4330;													
Best Local Similarity	99.7%;		Pred. No. 3.2e-157;																	
Matches	734;		Conservative		0;		Mismatches		1;		Indels									
QY	616	GGCTGCCCGCCTCACTGGGAAGCACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA	675																	
Db	3414	GGCTGCCCGCCTCACTGGGAAGCACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA	3473																	
QY	676	GAATTTTGGAGTGCACCAAGCTTTCTGTGAGACAACTCTTCACATCTTCTTTTCATA	735																	
Db	3474	GAATTTTGGAGTGCACCAAGCTTTCTGTGAGACAACTCTTCACATCTTCTTTTCATA	3533																	
QY	736	AACACTAGAGAGAACAGCAATGGATATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG	795																	
Db	3534	AACACTAGAGAGAACAGCAATGGATATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG	3593																	
QY	796	ATCGGCCTCACAGCTCAGAGCGTGAATGAATGAATGAATGCTGCTGGATGGGACATCTCCA	855																	
Db	3594	ATCGGCCTCACAGCTCAGAGCGTGAATGAATGAATGAATGCTGCTGGATGGGACATCTCCA	3653																	
QY	856	GACTACAAAAATTTGAAAGCTGGACAGCGGATTAAGTGGGGTCATGCGCCATGGCCAGGA	915																	
Db	3654	GACTACAAAAATTTGAAAGCTGGACAGCGGATTAAGTGGGGTCATGCGCCATGGCCAGGA	3713																	
QY	916	GAAGACTGTGCTGGGTGATTATGCTGGGCAGTGGAGCAATTTCCAATGTGAAGACGTC	975																	
Db	3714	GAAGACTGTGCTGGGTGATTATGCTGGGCAGTGGAGCAATTTCCAATGTGAAGACGTC	3773																	
QY	976	AATACTTCATTTGGAAAAAGACAGGGAGACAGTACTGTCATCTGCATTATACGGACT	1035																	
Db	3774	AATACTTCATTTGGAAAAAGACAGGGAGACAGTACTGTCATCTGCATTATACGGACT	3833																	
QY	1036	GTGATGGGATCATGAGCAAAATTTTCAGCTCTCAAAGCAAGGACACTCCCTTCTTAAT	1095																	
Db	3834	GTGATGGGATCATGAGCAAAATTTTCAGCTCTCAAAGCAAGGACACTCCCTTCTTAAT	3893																	
QY	1096	TGCATCACCTTCTCATCATGATTGAAAAAAGCACTGAAAAACCAATTAATCTGAAAAA	1155																	
Db	3894	TGCATCACCTTCTCATCATGATTGAAAAAAGCACTGAAAAACCAATTAATCTGAAAAA	3952																	
QY	1156	AAATTGACAGCTAGTGTGTTTTTACCATCGTCAATTACCCAAAGACTTGGGAATCAAAATG	1215																	
Db	3953	AAATTGACAGCTAGTGTGTTTTTACCATCGTCAATTACCCAAAGACTTGGGAATCAAAATG	4012																	
QY	1216	TTCCCCAGGTTATGCTGATTTTCATTTGCATGCGATGCAATCATCATGATTTCTCC	1275																	
Db	4013	TTCCCCAGGTTATGCTGATTTTCATTTGCATGCGATGCAATCATCATGATTTCTCC	4072																	
QY	1276	TCGGTCAGTAACCGTCGCAATTAACAAATATGTCTTCCAAAGTATGGAACACTCCAATC	1335																	
Db	4073	TCGGTCAGTAACCGTCGCAATTAACAAATATGTCTTCCAAAGTATGGAACACTCCAATC	4132																	
QY	1336	AGAAAAAGGTTATCAT	1351																	

Db	4133	AGAAAAAGGTTATCAT	4148
RESULT 10	AC016128/c		
LOCUS	AC016128.4		
DEFINITION	Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT SEQUENCE, 8 unordered pieces.		
ACCESSION	AC016128		
VERSION	AC016128.4		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 169088)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome 18, clone RP11-324G2		
REFERENCE	2 (bases 1 to 169088)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boquslavsky,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lebowitz,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testave,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-Nov-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Sep 9, 2000 this sequence version replaced gi:6649269. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html		
	----- Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Center code: WIBR		
	Web site: http://www-seq.wi.mit.edu		
	Contact: sequence_submissions@genome.wi.mit.edu		
	----- Project Information		
	Center project name: L999		
	Center clone name: 324.G.2		
	----- Summary Statistics		
	Sequencing vector: M13: M7815; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Assembly program: Phrap; version 0.960731		
	Consensus quality: 165522 bases at least Q40		
	Consensus quality: 167248 bases at least Q30		
	Consensus quality: 167996 bases at least Q20		
	Insert size: 177000; agarose-fp		
	Insert size: 168388; sum-of-contigs		
	Quality coverage: 6.2 in Q20 bases; agarose-fp		
	Quality coverage: 6.5 in Q20 bases; sum-of-contigs		

	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 8 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	* 1 17943: contig of 17943 bp in length		
	* 17944 18043: gap of 100 bp		
	* 18044 22424: contig of 4381 bp in length		

Db 4133 AAAAAAGGTTATCAT 4148

RESULT 10
AC016128/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
AC016128
AC016128 GI:10046526
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-324G2
Unpublished

2 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castelle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Leloczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wynan,D., Ye W.J., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
COMMENT
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L999
Center clone name: 324_G2
----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16522 bases at least Q40
Consensus quality: 167248 bases at least Q30
Consensus quality: 167996 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 168388; sum-of-ontigs
Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 bases; sum-of-ontigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17943: contig of 17943 bp in length
* 17944 18043: gap of 100 bp
* 18044 22424: contig of 4381 bp in length

* 22425 22524: gap of 100 bp
* 22525 38094: contig of 15570 bp in length
* 38095 38194: gap of 100 bp
* 38195 49220: contig of 11026 bp in length
* 49221 49320: gap of 100 bp
* 49321 66394: contig of 17074 bp in length
* 66395 66494: gap of 100 bp
* 66495 91692: contig of 25198 bp in length
* 91693 91792: gap of 100 bp
* 91793 120869: contig of 29077 bp in length
* 120870 120869: gap of 100 bp
* 120970 169088: contig of 48119 bp in length.

FEATURES

source
1. .169088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18"
/clone="RP11-324G2"
/clone_lib="RPCI-11 Human Male BAC"
1. .17943
/note="assembly_fragment
clone_end:SP6
vector_side:left"
18044. .22424
/note="assembly_fragment"
22525. .38094
/note="assembly_fragment"
38195. .49220
/note="assembly_fragment"
49321. .66394
/note="assembly_fragment"
66495. .91692
/note="assembly_fragment"
91793. .120869
/note="assembly_fragment"
120970. .169088
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN

Query Match 36.1%; Score 488.8; DB 2; Length 169088;
Best Local Similarity 99.6%; Pred. No. 9.2e-103;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 126 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGGGCC 185
Db 153817 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGGGCC 153758
QY 186 AACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCTGGACACCTGGCCCTGCGGG 245
Db 153757 AACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCTGGACACCTGGCCCTGCGGG 153698
QY 246 TGAGAGAGGCCCAATTGGACACGTGGTCCCGCCGGAGAGCTGGCGGCAAAAGATCTAA 305
Db 153697 TGAGAGAGGCCCAATTGGACACGTGGTCCCGCCGGAGAGCTGGCGGCAAAAGATCTAA 153638
QY 306 AGGCTCCAGGGCCCGCAAGAGCTCCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCC 365
Db 153637 AGGCTCCAGGGCCCGCAAGAGCTCCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCC 153578
QY 366 CAGTGGGAGCCAGCGCCCGCCCGCCACAGGCAAAAGAGGAGCTCCCGGCCCTCAGGG 425
Db 153577 CAGTGGGAGCCAGCGCCCGCCCGCCACAGGCAAAAGAGGAGCTCCCGGCCCTCAGGG 153518
QY 426 CCCTCCTGGCTTCCAGGAGCTTCAGGACACCGTTGGGAGAGCTGGGGTGCCTGGACCTCG 485
Db 153517 CCCTCCTGGCTTCCAGGAGCTTCAGGACACCGTTGGGAGAGCTGGGGTGCCTGGACCTCG 153458
QY 486 GGGACTGCGAGGCTTGCTGGGGTACAGGATGCCAGGCCCGCCCAAGGGCCCGCCCGGCC 545
|||||

Db 153457 GGGACTGCGAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCGCCCGCCCGGCC 153398
QY 546 TCCTGGCCCATCAGAGCGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGSCACC 605
|||||
Db 153397 TCCTGGCCCATCAGAGCGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGSCACC 153338
QY 606 GGAGGACAAATGG 617
|||||
Db 153337 GGAGGACAAATGG 153326
|||||
RESULT 11
AP000915/c 188439 bp DNA linear PRI 27-APR-2002
LOCUS Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete
DEFINITION sequences.
ACCESSION AP000915
VERSION AP000915.5 GI:20334314
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-720L2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 188439)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@isc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Apr 26, 2002 this sequence version replaced gi:9188470.
FEATURES
Location/Qualifiers
source
1..188439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p"
/clone="RP11-720L2"
BASE COUNT 54003 a 41134 c 40027 g 53275 t
ORIGIN
Query Match 36.1%; Score 488.8; DB 9; Length 188439;
Best Local Similarity 99.6%; Pred. No. 9.3e-103;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 126 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGGGCC 185
Db 44938 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGGGCC 44879
QY 186 AACTGGCAACAGGACAGAAAGAGAGAGGGGAGAGCTGGACACCTGGCCCTGCGGG 245
Db 44878 AACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGAGCTGGACACCTGGCCCTGCGGG 44819
QY 246 TGAGAGAGGCCCAATTGGACACGTGGTCCCGCCGGAGAGCGTGGCGGCAAAAGATCTAA 305
Db 44818 TGAGAGAGGCCCAATTGGACACGTGGTCCCGCCGGAGAGCGTGGCGGCAAAAGATCTAA 44759
QY 306 AGGCTCCAGGGCCCGCAAGAGCTCCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCC 365
Db 44758 AGGCTCCAGGGCCCGCAAGAGCTCCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCC 44699
QY 366 CAGTGGGAGCCAGCGCCCGCCCGCCACAGGCAAAAGAGGAGCTCCCGGCCCTCAGGG 425
Db 44698 CAGTGGGAGCCAGCGCCCGCCCGCCACAGGCAAAAGAGGAGCTCCCGGCCCTCAGGG 44639
QY 426 CCCTCCTGGCTTCCAGGAGCTTCAGGAGACCGTTGGGAGAGCTGGGGTGCCTGGACCTCG 485
|||||

```
|||||
Db 44638 CCTCTGCTTCCAGGACTTCAGGACACCGTTGGGAGCGTGGGGTCCCTGGACCTCG 44579
|||||
QY 486 GGGACTGCCAGGCTTGGCTGGGTTACCAAGGATGTCAGGCCCAAGGGCCCCCGGGCCC 545
|||||
Db 44578 GGGACTGCCAGGCTTGGCTGGGTTACCAAGGATGTCAGGCCCAAGGGCCCCCGGGCCC 44519
|||||
QY 546 TCCTGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCAACCCCGGCACC 605
|||||
Db 44518 TCCTGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCAACCCCGGCACC 44459
|||||
QY 606 GGAGGACATGG 617
Db 44458 GGAGGACATGG 44447

RESULT 12
AC024368 71044 bp DNA linear HTG 28-FEB-2000
LOCUS Homo sapiens chromosome 11 clone RP11-179K3 map 11, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC024368
VERSION AC024368.1 GI:7108157
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 71044)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McWan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6145
Center clone name: 179_K_3
-----
* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
```

```
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 698: contig of 698 bp in length
699 798: gap of 100 bp
799 1495: contig of 697 bp in length
1496 1595: gap of 100 bp
1596 2284: contig of 689 bp in length
2285 2384: gap of 100 bp
2385 3056: contig of 672 bp in length
3057 3156: gap of 100 bp
3157 3851: contig of 695 bp in length
3852 3951: gap of 100 bp
3952 4627: contig of 676 bp in length
4628 4727: gap of 100 bp
4728 5409: contig of 682 bp in length
5410 5509: gap of 100 bp
5510 6183: contig of 674 bp in length
6184 6283: gap of 100 bp
6284 6984: contig of 701 bp in length
6985 7084: gap of 100 bp
7085 7768: contig of 684 bp in length
7769 7868: gap of 100 bp
7869 8550: contig of 682 bp in length
8551 8650: gap of 100 bp
8651 9352: contig of 702 bp in length
9353 9452: gap of 100 bp
9453 10140: contig of 688 bp in length
10141 10240: gap of 100 bp
10241 10922: contig of 682 bp in length
10923 11022: gap of 100 bp
11023 11709: contig of 687 bp in length
11710 11809: gap of 100 bp
11810 12511: contig of 702 bp in length
12512 12611: gap of 100 bp
12612 13304: contig of 693 bp in length
13305 13404: gap of 100 bp
13405 14095: contig of 691 bp in length
14096 14195: gap of 100 bp
14196 14881: contig of 686 bp in length
14882 14981: gap of 100 bp
14982 15674: contig of 693 bp in length
15675 15774: gap of 100 bp
15775 16460: contig of 686 bp in length
16461 16560: gap of 100 bp
16561 17250: contig of 690 bp in length
17251 17350: gap of 100 bp
17351 18040: contig of 690 bp in length
18041 18140: gap of 100 bp
18141 18836: contig of 696 bp in length
18837 18936: gap of 100 bp
18937 19621: contig of 685 bp in length
19622 19721: gap of 100 bp
19722 20425: contig of 704 bp in length
20426 20525: gap of 100 bp
20526 21224: contig of 699 bp in length
21225 21324: gap of 100 bp
21325 22017: contig of 693 bp in length
22018 22117: gap of 100 bp
22118 22815: contig of 698 bp in length
22816 22915: gap of 100 bp
22916 23609: contig of 694 bp in length
23610 23709: gap of 100 bp
23710 24402: contig of 693 bp in length
24403 24502: gap of 100 bp
24503 25191: contig of 689 bp in length
25192 25291: gap of 100 bp
25292 25981: contig of 690 bp in length
25982 26081: gap of 100 bp
```

Query Match	35.8%	Score	484.6	DB	2	Length	71044
Best Local Similarity	98.2%	Prod. No.	8e-102				
Matches	490	Conservative	0	Mismatches	9	Indels	0
QY	119	TACTACAAGGTCCACCGGSCCCAGGGCTCCAAGAGGTGACAGAGGATCCACGAGCACCC	178				
Db	43599	TAGCAAAAGTCCACCGGSCCCAGGGCTCCAAGAGGTGACAGAGATCCACGAGCACCC	43658				
QY	179	CTGGCCCCAACTGGCAACAAGGACAGAAGAGAGAGAGGGGAGGCTTGACCAACCTGGCC	238				
Db	43659	CTGGCCCCAACTGGCAACAAGGACAGAAGAGAGAGAGGGGAGGCTTGACCAACCTGGCC	43718				
QY	239	CTGGGGGTGACAGAGGCCCAATTGGACAGACTGTGCCCCGGAGAGCGTGGCGGCAAG	298				
Db	43719	CTGGGGGTGACAGAGGCCCAAGTGTGCCCCGGAGAGCGTGGCGGCAAG	43778				
QY	299	GATCTAAAGGCTCCACAGGCCCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGGCCCTC	358				
Db	43779	GATCTAAAGGCTCCACAGGCCCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGGCCCTC	43838				
QY	359	AGGGCCCCAGTGGGAGACCCAGGCCCCCCGGGCCACCACGGAAGAGGAGTCTCCCCGGCC	418				
Db	43839	AGGGCTCCAGTGGGAGACCCAGGCCCCCCGGGCCACCACGGAAGAGGAGTCTCCCCGGCC	43898				
QY	419	CTCAGGGCCCTCTCGGCTTCCAGGACATTCAGGGACACCGTTGGGAGAGCCTGGGGTGCCCTG	478				
Db	43899	CTCAGGGCCCTCTCGGCTTCCAGGACATTCAGGGACACCGTTGGGAGAGCCTGGGGTGCCCTG	43958				
QY	479	GACCTCGGGGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGGGGCCCCC	538				
Db	43959	GACCTCGGGGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGGGGCCCCC	44018				
QY	539	CCGGCCCTCTCGGCCATCAGAGCGGTGTGCCCTTGCCCTGTCAGAAATGAGCCAAAGCC	598				
Db	44019	CCGGCCCTCTCGGCCATCAGAGCGGTGTGCCCTTGCCCTGTCAGAAATGAGCCAAAGCC	44078				
QY	599	CGGCACCGGAGACAAATGG	617				
Db	44079	CAGCACCGGAGACAAATGG	44097				

```

RESULT 13
AP001022      187635 bp      DNA      linear      HTG 30-MAY-2000
LOCUS      Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING
DEFINITION      DRAFT SEQUENCE, 33 unordered pieces.

ACCESSION      AP001022
VERSION      AP001022.2 GI:8117692
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-815L4.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 187635)
               Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
               Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
               Homo sapiens 187,635 genomic DNA of 18p11.3
               Published Only in Database (2000)
               2 (bases 1 to 187635)
               Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
               Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
               Direct Submission
               Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
               Japan (E-mail:hattori@gsc.riken.go.jp,
               URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
               Fax:81-42-778-9924)
               On May 31, 2000 this sequence version replaced gi:6997772.
               ----- Genome Center
               Center: RIKEN Genomic Sciences Center(GSC)
               Center code: RIKEN
               Web site: http://hgp.gsc.riken.go.jp/
COMMENT

```

Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-815L4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167192 bases at least Q40
Consensus quality: 177020 bases at least Q30
Consensus quality: 182105 bases at least Q20
Insert size: 184435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved
1 21332 contig of 21332 bp in length
21433 42072 contig of 20640 bp in length
42173 54336 contig of 12164 bp in length
54437 66766 contig of 12330 bp in length
66867 77849 contig of 10983 bp in length
77950 87798 contig of 9849 bp in length
87899 99510 contig of 11612 bp in length
99611 107273 contig of 100 bp
107374 13189 contig of 7663 bp in length
131290 120559 contig of 5816 bp in length
126569 13619 contig of 3015 bp in length
131720 144232 contig of 3381 bp in length
136911 156400 contig of 100 bp
140026 158933 contig of 2434 bp in length
14333 148902 contig of 100 bp
14333 148902 contig of 100 bp
148903 149002 contig of 100 bp
149003 152918 contig of 3916 bp in length
152919 153018 contig of 100 bp
153019 156399 contig of 3381 bp in length
156400 156499 contig of 100 bp
156500 158933 contig of 2434 bp in length
158934 159033 contig of 100 bp
159034 161884 contig of 2851 bp in length
161885 161984 contig of 100 bp
161985 164301 contig of 2317 bp in length
164302 164401 contig of 100 bp
164402 166300 contig of 1899 bp in length
166301 166400 contig of 100 bp
166401 168580 contig of 2180 bp in length
168581 168680 contig of 100 bp
168681 171270 contig of 2590 bp in length
171271 171370 contig of 100 bp
171371 174092 contig of 2722 bp in length
174093 174192 contig of 100 bp
174193 176095 contig of 1903 bp in length
176096 176195 contig of 100 bp
176196 177890 contig of 1695 bp in length
177891 177990 contig of 100 bp
177991 179786 contig of 1796 bp in length
179787 179886 contig of 100 bp
179887 181478 contig of 1592 bp in length
181479 181579 contig of 100 bp
181579 182895 contig of 1317 bp in length
182896 182996 contig of 100 bp
182996 184841 contig of 1846 bp in length
184842 184941 contig of 100 bp
184942 186345 contig of 1404 bp in length
186346 186445 contig of 100 bp
186446 187635 contig of 1190 bp in length.
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 21332: contig of 21332 bp in length
21433 21432: gap of 100 bp
21433 42072: contig of 20640 bp in length
42073 42172: gap of 100 bp
42173 54336: contig of 12164 bp in length
54337 54436: gap of 100 bp
54437 66766: contig of 12330 bp in length
66767 66866: gap of 100 bp
66867 77849: contig of 10983 bp in length
77850 77949: gap of 100 bp

77950 87798: contig of 9849 bp in length
87799 87898: gap of 100 bp
87899 99510: contig of 11612 bp in length
99511 99610: gap of 100 bp
99611 107273: contig of 7663 bp in length
107274 107373: gap of 100 bp
107374 13189: contig of 5816 bp in length
13189 13190: gap of 100 bp
13190 120559: contig of 7270 bp in length
120560 120659: gap of 100 bp
120660 128468: contig of 5809 bp in length
128469 128568: gap of 100 bp
128569 131619: contig of 5051 bp in length
131620 131719: gap of 100 bp
131720 136810: contig of 5091 bp in length
136811 136910: gap of 100 bp
136911 139925: contig of 3015 bp in length
139926 140025: gap of 100 bp
140026 144232: contig of 4207 bp in length
144233 144332: gap of 100 bp
144333 148902: contig of 4570 bp in length
148903 149002: gap of 100 bp
149003 152918: contig of 3916 bp in length
152919 153018: gap of 100 bp
153019 156399: contig of 3381 bp in length
156400 156499: gap of 100 bp
156500 158933: contig of 2434 bp in length
158934 159033: gap of 100 bp
159034 161884: contig of 2851 bp in length
161885 161984: gap of 100 bp
161985 164301: contig of 2317 bp in length
164302 164401: gap of 100 bp
164402 166300: contig of 1899 bp in length
166301 166400: gap of 100 bp
166401 168580: contig of 2180 bp in length
168581 168680: gap of 100 bp
168681 171270: contig of 2590 bp in length
171271 171370: gap of 100 bp
171371 174092: contig of 2722 bp in length
174093 174192: gap of 100 bp
174193 176095: contig of 1903 bp in length
176096 176195: gap of 100 bp
176196 177890: contig of 1695 bp in length
177891 177990: gap of 100 bp
177991 179786: contig of 1796 bp in length
179787 179886: gap of 100 bp
179887 181478: contig of 1592 bp in length
181479 181578: gap of 100 bp
181579 182895: contig of 1317 bp in length
182896 182995: gap of 100 bp
182996 184841: contig of 1846 bp in length
184842 184941: gap of 100 bp
184942 186345: contig of 1404 bp in length
186346 186445: gap of 100 bp
186446 187635: contig of 1190 bp in length.
Location/Qualifiers
1. 187635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-815L4"
1. 21332
/note="assembly_fragment"
21433. 42072
/note="assembly_fragment"
42173. 54336
/note="assembly_fragment"
54437. 66766
/note="assembly_fragment"
66867. 77849
/note="assembly_fragment"
77950. 87798
FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens DNA, clone:RP11-683J11.
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178022)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 178,022 genomic DNA of 18p11.3
Published Only in DataBase (1999)

```

web site: http://ngp.gs.riken.go.jp/
Contact: hattori@gs.riken.go.jp
----- Project Information -----
Center project name: HumDraft18
Center clone name: Rp11-683J11
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of
Assembly program: Phrap; version 0.930329
Consensus quality: 161944 bases at least Q40
Consensus quality: 170357 bases at least Q30
Consensus quality: 174322 bases at least Q20
Insert size: 176122; sum-of-contigs
Quality coverage: 4.56x in Q20 bases sum-of-c

```

Preserved	1	24283	contig of	24283	bp	in length
	24384	46875	contig of	24292	bp <td>in length</td>	in length
	46976	65058	contig of	18083	bp <td>in length</td>	in length
	65159	81322	contig of	16164	bp <td>in length</td>	in length
	81423	98823	contig of	17401	bp <td>in length</td>	in length
	98924	110068	contig of	11445	bp <td>in length</td>	in length
	110169	122716	contig of	12548	bp <td>in length</td>	in length
	122817	133089	contig of	10273	bp <td>in length</td>	in length
	133190	141206	contig of	8017	bp <td>in length</td>	in length
	141307	149050	contig of	7744	bp <td>in length</td>	in length
	149151	156105	contig of	6955	bp <td>in length</td>	in length
	156206	162048	contig of	5843	bp <td>in length</td>	in length
	162149	165118	contig of	2970	bp <td>in length</td>	in length
	165219	167480	contig of	2262	bp <td>in length</td>	in length
	167581	169986	contig of	2406	bp <td>in length</td>	in length
	170087	171909	contig of	1823	bp <td>in length</td>	in length
	172010	173217	contig of	308	bp <td>in length</td>	in length
	172418	174663	contig of	2246	bp <td>in length</td>	in length
	174764	176271	contig of	1508	bp <td>in length</td>	in length
	176372	178022	contig of	1651	bp <td>in length</td>	in length

170372 170022 Conting 01 1051 bp in length
Sequence updated (24-Dec-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* gaps of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 24283: contig of 24283 bp in length
24284 24383: gap of 100 bp
24384 46875: contig of 22492 bp in length
46876 46975: gap of 100 bp
46976 65058: contig of 18083 bp in length
65059 65158: gap of 100 bp
65159 81322: contig of 16164 bp in length
81323 81422: gap of 100 bp
81423 98823: contig of 17401 bp in length
98824 98923: gap of 100 bp
98924 110068: contig of 11145 bp in length
110069 110168: gap of 100 bp
110169 122716: contig of 12548 bp in length
122717 122816: gap of 100 bp
122817 133089: contig of 10273 bp in length
133090 133189: gap of 100 bp
133190 141206: contig of 8017 bp in length
141207 141306: gap of 100 bp
141307 149050: contig of 7744 bp in length
149051 149150: gap of 100 bp
149151 156105: contig of 6955 bp in length
156106 156205: gap of 100 bp
156206 162048: contig of 5843 bp in length
162049 162148: gap of 100 bp
162149 165118: contig of 2970 bp in length
165119 165218: gap of 100 bp
165219 167480: contig of 2262 bp in length
167481 167580: gap of 100 bp
167581 169986: contig of 2406 bp in length
169987 170086: gap of 100 bp
170087 171909: contig of 1823 bp in length
171910 172009: gap of 100 bp
172010 172317: contig of 308 bp in length
172318 172417: gap of 100 bp
172418 174663: contig of 2246 bp in length
174664 174763: gap of 100 bp
174764 176271: contig of 1508 bp in length
176272 176371: gap of 100 bp
176372 178022: contig of 1651 bp in length.

FEATURES

source
1. .178022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-683J11"
1. .24283
/note="assembly_fragment"
24384. .46875
/note="assembly_fragment"
46976. .65058
/note="assembly_fragment"
65159. .81322
/note="assembly_fragment"
81423. .98823
/note="assembly_fragment"
98924. .110068
/note="assembly_fragment"
110169. .122716
/note="assembly_fragment"
122817. .133089
/note="assembly_fragment"
133190. .141206
/note="assembly_fragment"
141307. .149050
/note="assembly_fragment"
149151. .156105
/note="assembly_fragment"
156206. .162048
/note="assembly_fragment"

misc_feature 162149. .165118
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 165219. .167480
/note="assembly_fragment"
misc_feature 167581. .169986
/note="assembly_fragment"
misc_feature 170087. .171909
/note="assembly_fragment"
misc_feature 172010. .172317
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 172418. .174663
/note="assembly_fragment"
misc_feature 174764. .176271
/note="assembly_fragment"
misc_feature 176372. .178022
/note="assembly_fragment"
BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others
ORIGIN

Query Match 34.4%; Score 465.8; DB 2; Length 178022;
Best Local Similarity 99.2%; Pred. No. 2e-97;
Matches 489; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 126 AGGTCACCGGGCCCCAGGGGTCCCAAGAGGTCACAGAGATCCCAGGGACCCCTGGGCC 185
|||||
Db 123478 AGGTCACCGGGCCCCAGGGGTCCCAAGAGGTGACAGAGATCCCAGGGACCCCTGGGCC 123537
QY 186 AACTGGCAACAAGGACAGAAAGGAGAGAGAGGGGAGCCTGGACCACCTGGGCCCTGGCGG 245
|||||
Db 123538 AACTGGCAACAAGGACAGAAAGGAGAGAGAGGGGAGCCTGGACCACCTGGGCCCTGGCGG 123597
QY 246 TGAGAGAGGCCCAATGGACCACTGGTCCCTCCCGGAGAGCCTGGCGGCAAGGATCTAA 305
Db 123598 TGAGAGAGGCCCAATGGACCACTGGTCCCTCCCGGAGAGCCTGGCGGCAAGGATCTAA 123657
QY 306 AGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCC 365
|||||
Db 123658 AGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCTC 123717
QY 366 CAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGGCCCTCAGGG 425
Db 123718 CAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGGCCCTCAGGG 123777
QY 426 CCTCTGGCTTCCAGGGACTTCAGGGACCTGTTGGGAGCCTGGGGTGCCTGGACCTCG 485
|||||
Db 123778 CCTCTGGCTTCCAGGGACTTCAGGGACCTGTTGGGAGCCTGGGGTGCCTGGACCTCG 123837
QY 486 GGGACTGCCAGGCTTGCCTGGGGTA-CCAGGCATGCCAGGGCCCCAAGGGCCCCCGGCC 544
Db 123838 GGGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGGCCCCAAGGGCCCCCGGCC 123897
QY 545 CTCCTGGCCCATCAGAGCGGTGGTCCCTGGCCCTGGCCCTGCAGAAATGAGCAACCCCGGCAC 604
Db 123898 CTCCTGGCCCATCAGAGCGGTGGTG-CCCTGGCCCTGCAGAAATGAGCAACCCCGGCAC 123956
QY 605 CGAGGACAATGG 617
Db 123957 CGAGGACAATGG 123969

RESULT 15

AP000939/c
LOCUS AP000939 18255 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP000939
VERSION AP000939.3 GI:8119080
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-839023.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 18255)

AUTHORS

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 188,255 genomic DNA of 18p11.3
Published Only in DataBase (1999)
2 (bases 1 to 188255)

AUTHORS

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission

TITLE

Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail: hattori@gsc.riken.go.jp)
URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9923,
Fax: 81-42-778-9924

COMMENT

On May 31, 2000 this sequence version replaced gi:6997751.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-839023

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 171068 bases at least Q40

Consensus quality: 178865 bases at least Q30

Consensus quality: 182445 bases at least Q20

Insert size: 185155; sum-of-contigs

Quality coverage: 4.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 16096 contig of 16096 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 63558 contig of 12448 bp in length
63659 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 6072 bp in length
91010 98615 contig of 7606 bp in length
98716 105978 contig of 7263 bp in length
106079 113049 contig of 5817 bp in length
113150 119201 contig of 6052 bp in length
119302 126582 contig of 7281 bp in length
126683 132070 contig of 5388 bp in length
132171 137138 contig of 4968 bp in length
137239 143055 contig of 5817 bp in length
143156 148639 contig of 5484 bp in length
148740 153086 contig of 4347 bp in length
153187 158921 contig of 4153 bp in length
158922 163174 contig of 3792 bp in length
163175 167066 contig of 3924 bp in length
167067 170120 contig of 2954 bp in length
170121 170220 contig of 100 bp in length
170221 172430 contig of 2210 bp in length
172431 172530 contig of 100 bp in length
172531 175598 contig of 3068 bp in length
175599 177118 contig of 1420 bp in length
177119 177218 contig of 100 bp in length
177219 178531 contig of 1313 bp in length
178532 178631 contig of 100 bp in length
178632 179857 contig of 1226 bp in length
179858 179957 contig of 1107 bp in length
179959 181064 contig of 1107 bp in length
181065 181164 contig of 100 bp in length
181165 182735 contig of 1571 bp in length
182736 182855 contig of 100 bp in length
182856 183923 contig of 1088 bp in length
183924 184023 contig of 100 bp in length
184024 185810 contig of 1787 bp in length
185811 185910 contig of 100 bp in length
185911 187145 contig of 1235 bp in length
187146 187245 contig of 1010 bp in length
187246 188255 contig of 1010 bp in length

Sequence updated (06-Jan-2000)

FEATURES

Location/Qualifiers

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 16096: contig of 16096 bp in length
* 16097 16196: gap of 100 bp
* 16197 33789: contig of 17593 bp in length
* 33790 33889: gap of 100 bp
* 33890 51010: contig of 17121 bp in length
* 51011 51110: gap of 100 bp
* 51111 63558: contig of 12448 bp in length
* 63559 63658: gap of 100 bp
* 63659 75056: contig of 11398 bp in length
* 75057 75156: gap of 100 bp
* 75157 84737: contig of 9581 bp in length
* 84738 84837: gap of 100 bp
* 84838 90909: contig of 6072 bp in length
* 90910 91009: gap of 100 bp
* 91010 98615: contig of 7606 bp in length
* 98616 98715: gap of 100 bp
* 98716 105978: contig of 7263 bp in length
* 105979 106078: gap of 100 bp
* 106079 113049: contig of 6971 bp in length
* 113050 113149: gap of 100 bp
* 113150 119201: contig of 6052 bp in length
* 119202 119301: gap of 100 bp
* 119302 126582: contig of 7281 bp in length
* 126583 126682: gap of 100 bp
* 126683 132070: contig of 5388 bp in length
* 132071 132170: gap of 100 bp
* 132171 137138: contig of 4968 bp in length
* 137139 137238: gap of 100 bp
* 137239 143055: contig of 5817 bp in length
* 143056 143155: gap of 100 bp
* 143156 148639: contig of 5484 bp in length
* 148640 148739: gap of 100 bp
* 148740 153086: contig of 4347 bp in length
* 153087 153186: gap of 100 bp
* 153187 158921: contig of 5735 bp in length
* 158922 159021: gap of 100 bp
* 159022 163174: contig of 4153 bp in length
* 163175 163274: gap of 100 bp
* 163275 167066: contig of 3792 bp in length
* 167067 167166: gap of 100 bp
* 167167 170120: contig of 2954 bp in length
* 170121 170220: gap of 100 bp
* 170221 172430: contig of 2210 bp in length
* 172431 172530: gap of 100 bp
* 172531 175598: contig of 3068 bp in length
* 175599 175698: gap of 100 bp
* 175699 177118: contig of 1420 bp in length
* 177119 177218: gap of 100 bp
* 177219 178531: contig of 1313 bp in length
* 178532 178631: gap of 100 bp
* 178632 179857: contig of 1226 bp in length
* 179858 179957: gap of 100 bp
* 179959 181064: contig of 1107 bp in length
* 181065 181164: gap of 100 bp
* 181165 182735: contig of 1571 bp in length
* 182736 182855: gap of 100 bp
* 182856 183923: contig of 1088 bp in length
* 183924 184023: gap of 100 bp
* 184024 185810: contig of 1787 bp in length
* 185811 185910: gap of 100 bp
* 185911 187145: contig of 1235 bp in length
* 187146 187245: gap of 100 bp
* 187246 188255: contig of 1010 bp in length.


```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-635-086-1

Query Match      8.4%; Score 113.6; DB 1; Length 3181;
Best Local Similarity 52.5%; Pred. No. 7.6e-20;
Matches 248; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 97 CAGCTCATCAGAAATTTTACAACTACAGGTCCACCGGCCCCCAGGGGTCCAGAGGT 156
Db 104 CTGGTCCCAAGGCTTCCAAAGGCTTCCAAAGTCCCTCCCTGAGCCTGCGGAGCCTGGA 163

QY 157 GACAGAGGATCCAGGACCCCTCGGCCCAACTGCAACAAGGGACAGAAAGGAGAGAAG 216
Db 164 GCTTCAGTCCGANTGGTCCCGAGGTCCCGAGGTCCCGAGTCCCTGGAAAGATGGAGATGAT 223

QY 217 GGGGAGCTTGGACACCTTGGCCCTCGGGTGGAGAGAGGCCCAATTTGACCACTTGGTCCC 276
Db 224 GGGGAAGCTGGAACCTTGGTCTGCTGAGCGTGGGCCCTTCTTGGCCCTCAGGGTGT 283

QY 277 CCCGAGAGAGCTGGCGGCAAGGATCTAAAGCTCCCGAGGGCCCCAAAGGCTCCCGTGGT 336
Db 284 CGAGGATTTGCCCGGAACAGCTGGCCCTCGCTGGAATGAAGGGACACAGAGGTTTCAAGTGT 343

QY 337 TCCCTCGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCCCGGGCCCCACA 396
Db 344 TTGATGTGCCAAGGAGATGCTGGTCTGCTGTCTTAAGGGTGAGCCTGGCAGCCCT 403

QY 397 GCGAAGAGGACTCCCGGCCCTCAGGGCCCTCTGCTTCCAGGGACTTCCAGGGCACC 456
Db 404 GGTGAATGAGGACTCTGTTGAGTGGCCCGCCCGTGGCCCTGCTGGTGGAGAGGTGCG 463

QY 457 GTTGGGAGCTTGGGTGCTGGACTCGGGACTCGGGAGCTGCGCAGGCTTGCCTGGGGTACCAAGC 516
Db 464 CTGGAGGCCCTTGGCCCTTGGTGTGCTGTGGAATGATGTTGCTACTTGGTGTGCTGCCGG 523

QY 517 ATGCAGGCCCCAGGGCCCCCGGCCCTCTCTGCCCATCAGGAGCGGTGG 568
Db 524 CCCCCTGTCCACAGCGGCCCGCTGGTCTCTCTGCTTCCCTGGTGTGCTGTG 575

RESULT 4
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESS: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
```

```
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55536-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-48

Query Match      7.8%; Score 106; DB 1; Length 432;
Best Local Similarity 54.0%; Pred. No. 3e-18;
Matches 217; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 163 GGATCCCAAGGACCCCTCGGCCAACTGCAACAAGGGACAGAAAGGAGAGGGGGAG 222
Db 1 GGTGCTCAGCGCCACAGCAGGTCCCGAAGGGCGCATGGCCACAGAGGCCCGGAAAGGTGG 60

QY 223 CCTGACACACCTGGCCCTCGCGGGTGAGAGAGGCCCAATTTGACCACTGGTCCCCCGGA 282
Db 61 CTTGACCCGGTGTGTCACCGGGTGTCTCGGGACCTCGAGGCCCGCCAGGTGCGCTCGA 120

QY 283 GAGCCTGGCGCAAGGATCTAAAGCTCCCGAGGGCCCCAAAGGCTCCCGTGGTTCCT 342
Db 121 CGGGTGTGCTCACCAGGGTGTCTCCGGGACCTGCGAGGCCCGCCAGGTGCGCTTGAC 180

QY 343 GGGAAAGCCCGCCCTCAGGGCCCGAGTGGGGACCCAGGCCCGCCCGGGCCACAGGCAAA 402
Db 181 GGTCCACCGGGTGTCTCCGGGACCTCGAGGCCCGCCAGGTGCGCTTGAGCCGCTGGTCCA 240

QY 403 GAGGACTCCCGGCCCTCAGGGCCCTCTGCTTCCAGGGACTTCCAGGGCACCCTTGGG 462
Db 241 CCGGTGTCTCCGGGACCTGCGAGGCCCGCGAGGTGCGCTTGAGCCGCTGTCACCGGT 300

QY 463 GAGCCTGGGGTGTCTGGACCTGGGACTGCCAGGCTTCCAGGCTGCTCCGGGTACCATGCCA 522
Db 301 GCTCCGGGACCTGCGAGGCCCGCGAGGTGCGCTTGACCGGTGGTCCACCGGTGCTCCG 360

QY 523 GGCCCAAGGGCCCCCGGCCCTCTCTGSCCATCAGGAGCG 564
Db 361 GGACCTGCGAGGCCCGCCAGGTGCTCCATGGCCACAGGACCG 402

RESULT 5
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-19

Query Match      7.8%; Score 106; DB 4; Length 1608;
Best Local Similarity 52.8%; Pred. No. 5.2e-18;
Matches 229; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
```

```

QY 127 GGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCAGGGACCCCTCGGCCCA 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GGTGAACCTGGTCAAACTGGTCTCGAGTGCACCTGGTCTCATGGCCCGTGGTCT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 187 ACTGCAACAGGACACAAAGGAGAGAGGGGAGGCTGGACCACTGGCCCTCGGGT 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 GGTGCAACACATGGAACCGTGGTGAACCTTCTGGTCTCTTCTGGTCTCTGCTGT 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 GAGAGAGGCCCAATTTGGACAGCTGTGCCCGCGGAGAGCGTGGCGCAAGATCTAAA 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 GCTGTGGCCCAAGAGTCTTAGTGGCCACAGACATCTGTGGCGATAAGGAGAGCCC 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 GGTCTCCAGGGCCCCAAAGGCTTCCCGTGGTTCCTCCCTGGGAAGCCCGCCCTCAGGGCCCC 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GGTGAAAGGGGCCAGAGGTCTTCTGGCTTCAAGGGACACAATGATGGAAGTCTG 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 AGTGGGACCCAGGCCCCCGGGCCCCCAGGCAAGAGAGGACTCCCGGGCCCTCAGGGC 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 CCGTGTATCGTGGTGCACATGGTGATCAAGGTGCTCCTGGTCTCGTCTGTGT 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 CCGTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGTGCTGTGACCTCGG 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 CTAAGGGCCCTGCTGCTTCTTCTGGCCCTGCTGGAAGATGGTGCACACTGCATCT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 487 GGAATCCAGGCTTCCCTGGGTACCAAGCATGACAGGCCCCCAAGGGCCCCCGGGCCCT 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GGTAGCGTTGGACCTGCTGGCATTCGAGGCCCTCAGGGTCAACCAAGGCCCTGCTGCCCC 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 547 CCGTGGCCCATCAGG 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 CCGTGGTCCCCCTGG 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

```

US-08-494-168-1
; Sequence 1, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeaders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5102 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(2..82, 86..97, 101..4399, 4403..4420, 4424
; LOCATION: ..4465, 4469..4876, 4880..5101)
; US-08-494-168-1
Query Match
Best Local Similarity 7.4%; Score 100; DB 1; Length 5102;
Matches 220; Conservativity 52.4%; Pred. No. 3e-16;
Mismatch 0; Mismatch 200; Indels 0; Gaps 0;
QY 123 ACAAGGTCCACGGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGGACCCCTGG 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2380 RAATGGTCTCGGGGGAACAAGGCTTACAAGGATTACAGGGCACAAAGGATTTCTTGG 2439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 CCCAACTGGCAACAAGGACACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTTGGCCCTGC 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2440 AGACTCTGTGGCTTCCAGGACTCAAGGGTGTGACGGGAAGCTTACTAGGCCCA 2499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 243 GGTGAGAGAGGCCCAATTTGGACAGCTGTGCCCGGAGAGAGCTGGCGCAAGGATC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2500 AGGTGAGCGGGCACCCCTGGGACACAGGACAGGTGGGACAGCCAGGCCACCCAGGATC 2559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 303 TAAAGGCTCCAGGGCCCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCAGGG 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2560 TAGTGTCCATATGTCATCAAGGGCAAAATCTGGGCTCCAGGAGCACCAAGGCTTCCAGG 2619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 363 CCCAGTGGGACCCAGGCCCCCGGGCCCCCAGGCAAGAGAGGACTCCCGGCCCTCA 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2620 CATCTCAGGACATCTCTGGAAGAAAGAACAAAGAGTCTCTCTGGATCAAT 2679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 423 GGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGCTGGGGAGCCCTGGGGTGCCTGGACC 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2680 TGTAAGAAAGGGCTGCCAGGCTTAAAGGCCCTTCTGGAATCAGGCTAGTAGGACT 2739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 483 TCGGGACTGCCAGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGG 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2740 GAAAGAACCCAGGCTCTCCAGGGGTGCTGGTGGTCCAGGCCCTCTCTGGACCCCAAGGG 2799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 7

```

US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeaders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621,091G
; FILING DATE: 11/30/90
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA: No. 5424408 applicable
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard S.
; REGISTRATION NUMBER: 28180
; REFERENCE/DOCKET NUMBER: 900983/RB

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: No. 5424408 known
CHROMOSOME/SEGMENT: No. 5424408 known
PUBLICATION INFORMATION: No. 5424408e
US-07-621-091G-1

Query Match 7.4%; Score 99.6; DB 1; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2.2e-16;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
QY 127 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
DB 127 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
QY 283 GGAGCACCAGGTGTCCCGGTGACAGGGCCAGAGGTGATCTTGATTTATGATTT 342
DB 283 GGAGCACCAGGTGTCCCGGTGACAGGGCCAGAGGTGATCTTGATTTATGATTT 342
QY 187 ACTGCGACACAGGGACAGAAAGGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCTCGGGT 246
DB 187 ACTGCGACACAGGGACAGAAAGGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCTCGGGT 246
QY 343 CCAGGCATGAAAGGGAAGAGGGTAATTCAGGATTTCCAGGACCCACCTGGACCTCCAGGG 402
DB 343 CCAGGCATGAAAGGGAAGAGGGTAATTCAGGATTTCCAGGACCCACCTGGACCTCCAGGG 402
QY 247 GAGAGAGGCCCAATTTGGACCACTGGTCCCGGCGAGAGCGGTGGGGGCAAGAGGATCTAAA 306
DB 247 GAGAGAGGCCCAATTTGGACCACTGGTCCCGGCGAGAGCGGTGGGGGCAAGAGGATCTAAA 306
QY 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCCTGGCACAGTGAAGATCATC 462
DB 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCCTGGCACAGTGAAGATCATC 462
QY 307 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
DB 307 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
QY 463 TCCCTTCCAGGAGAGCCAGGCCCCACCTGGTTCAGCTGGAGAACACCGGATGCAAGAGAA 522
DB 463 TCCCTTCCAGGAGAGCCAGGCCCCACCTGGTTCAGCTGGAGAACACCGGATGCAAGAGAA 522
QY 367 AGTGGGAGCCCGAGGCCCCCGGGCCCCACAGGCAAGAGAGGACTCCCGGCGCCCTCAGGGC 426
DB 367 AGTGGGAGCCCGAGGCCCCCGGGCCCCACAGGCAAGAGAGGACTCCCGGCGCCCTCAGGGC 426
QY 523 CCCGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACACAGGG 582
DB 523 CCCGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACACAGGG 582
QY 427 CCTCTGGCTTCCAGGGACTTCAGGSCACCTTCAGGSCACCTTCAGGSCACCTTCAGGSC 486
DB 427 CCTCTGGCTTCCAGGGACTTCAGGSCACCTTCAGGSCACCTTCAGGSCACCTTCAGGSC 486
QY 583 GAGGATGTGTCACCAAGAACTCTGGACCACTGGAGAAAAGGCAACAAAGTTGTAAA 642
DB 583 GAGGATGTGTCACCAAGAACTCTGGACCACTGGAGAAAAGGCAACAAAGTTGTAAA 642
QY 487 GGACTGCCAGGCTTGCCTGGGTACCCAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCT 546
DB 487 GGACTGCCAGGCTTGCCTGGGTACCCAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCT 546
QY 643 GGAGAGCAAGGACCACTGGATCCGATGCGCTGCCAGGCTTGAAGGGGAAACCTGGAGAC 702
DB 643 GGAGAGCAAGGACCACTGGATCCGATGCGCTGCCAGGCTTGAAGGGGAAACCTGGAGAC 702
QY 547 CTGCGCCCATCAGGAGCGGTGG 568
DB 547 CTGCGCCCATCAGGAGCGGTGG 568
QY 703 ACTGGACCACCTGCAGCAGGGG 724
DB 703 ACTGGACCACCTGCAGCAGGGG 724

RESULT 8
US-08-399-889-1
Sequence 1, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Calf
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1416)
US-08-399-889-1
Query Match 7.4%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2.2e-16;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
QY 127 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
DB 127 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
QY 283 GGAGCACCAGGTGTCCCGGTGACAGGGCCAGAGGTGATCTTGATTTATGATTT 342
DB 283 GGAGCACCAGGTGTCCCGGTGACAGGGCCAGAGGTGATCTTGATTTATGATTT 342
QY 187 ACTGCGACACAGGGACAGAAAGGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCTCGGGT 246
DB 187 ACTGCGACACAGGGACAGAAAGGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCTCGGGT 246
QY 343 CCAGGCATGAAAGGGAAGAGGGTAATTCAGGATTTCCAGGACCCACCTGGACCTCCAGGG 402
DB 343 CCAGGCATGAAAGGGAAGAGGGTAATTCAGGATTTCCAGGACCCACCTGGACCTCCAGGG 402
QY 247 GAGAGAGGCCCAATTTGGACCACTGGTCCCGGCGAGAGCGGTGGGGGCAAGAGGATCTAAA 306
DB 247 GAGAGAGGCCCAATTTGGACCACTGGTCCCGGCGAGAGCGGTGGGGGCAAGAGGATCTAAA 306
QY 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCCTGGCACAGTGAAGATCATC 462
DB 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCCTGGCACAGTGAAGATCATC 462
QY 307 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
DB 307 GGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGGTCCAGGAGTCCAGGGACCCCTGGCCCCA 186
QY 463 TCCCTTCCAGGAGAGCCAGGCCCCACCTGGTTCAGCTGGAGAACACCGGATGCAAGAGAA 522
DB 463 TCCCTTCCAGGAGAGCCAGGCCCCACCTGGTTCAGCTGGAGAACACCGGATGCAAGAGAA 522
QY 367 AGTGGGAGCCCGAGGCCCCCGGGCCCCACAGGCAAGAGAGGACTCCCGGCGCCCTCAGGGC 426
DB 367 AGTGGGAGCCCGAGGCCCCCGGGCCCCACAGGCAAGAGAGGACTCCCGGCGCCCTCAGGGC 426
QY 523 CCCGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACACAGGG 582
DB 523 CCCGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACACAGGG 582
QY 427 CCTCTGGCTTCCAGGGACTTCAGGSCACCTTCAGGSCACCTTCAGGSCACCTTCAGGSC 486
DB 427 CCTCTGGCTTCCAGGGACTTCAGGSCACCTTCAGGSCACCTTCAGGSCACCTTCAGGSC 486
QY 583 GAGGATGTGTCACCAAGAACTCTGGACCACTGGAGAAAAGGCAACAAAGTTGTAAA 642
DB 583 GAGGATGTGTCACCAAGAACTCTGGACCACTGGAGAAAAGGCAACAAAGTTGTAAA 642
QY 487 GGACTGCCAGGCTTGCCTGGGTACCCAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCT 546
DB 487 GGACTGCCAGGCTTGCCTGGGTACCCAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCT 546
QY 643 GGAGAGCAAGGACCACTGGATCCGATGCGCTGCCAGGCTTGAAGGGGAAACCTGGAGAC 702
DB 643 GGAGAGCAAGGACCACTGGATCCGATGCGCTGCCAGGCTTGAAGGGGAAACCTGGAGAC 702
QY 547 CTGCGCCCATCAGGAGCGGTGG 568
DB 547 CTGCGCCCATCAGGAGCGGTGG 568
QY 703 ACTGGACCACCTGCAGCAGGGG 724
DB 703 ACTGGACCACCTGCAGCAGGGG 724
RESULT 9
US-09-167-364-1
Sequence 1, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Calf


```
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 5691197 relevant
; ANTI-SENSE: No. 5691197 relevant
US-08-392-367B-1

Query Match          7.3%; Score 99.4; DB 1; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.8e-16;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGTCCACCGGGCCCGCCAGAGGTCCAAAGAGGTGACAGAGATCCACAGGACCCCTGGC 183
    ||||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 721 CCAGGTGTCCAAGGACCCCGCCAGGCGCCACAGGAGGAGGAGGAGGAGGAGGAGGAG 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 184 CCAACTGSCAACAAAGGACAAAGGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 781 CTTACGGGTGCACAGGGAACAGGAGCAACTGTGTCTCCAGGACCTCGAGGAGAGAAG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 244 GGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCGCCGAGAGCGTGGCGGCAAGGATCT 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 841 GGCAGCAAGGTGACATAGGTCTCAGTGGCCCAAGGGGGAACATGGCCACCAAGGGAGAC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 304 AAAGCTCCAGGGGCCCAAGAGTCCCGTGTTCCTCCCTGGGAAGCCCGCCCTCAGGGC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 901 AAAGGGGACCTAGGCTTCCAGGAAACAAAGGGGACATGGGCATGAAGGAGACACAGGGG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 364 CCCAGTGGGACCCAGGCGCCCGCCAGGACCCAGGAGAGGAGGAGGAGGAGGAGGAGGAG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 961 CCCATGGGGTCCCTGGAGCTCAGGGAGGTAAAGGTATGTGTGAAACACCGGCTTACCA 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 424 GGCCTCTCTGGCTTCCAGGGACTTCAGGCGACCCGTTGGGGAGCCCTGGGGTGGCTGGACCT 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1021 GGTGGTGGTGGATCTCCAGGAGTCAAGGTGACCAAGGAAACCTGGAGTCAGGGTGT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 484 CGGGAGTCCAGGCTTCCTGGGGTACAGGCATGCCAGGCCCAAGGGCCCGCCCGCCG 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1081 CCAGGCCCTCAAGGTGACAGGACTTTTCAGGTGCCAAGGGTGGAGCCAGGAGGAGGAG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 544 CTCTCTGGCCCATCAGGAGCGGTGG 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1141 CTTCTGGGCCAGGAGACCCCGG 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 12
US-08-893-467A-1
; Sequence 1, Application US/08893467A
; Patent No. 6063901
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Elomaa, Outi
; APPLICANT: Kangas, Maarit
; TITLE OF INVENTION: An Isolated DNA Sequence For a
; Patent No. 6063901
; TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
; TITLE OF INVENTION: a Collagenous Domain and the
; TITLE OF INVENTION: Polypeptide Chain Encoded by
; TITLE OF INVENTION: such a Sequence
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 720 Kb storable
; COMPUTER: IBM PS/2, Model 35 SX
; OPERATING SYSTEM: DOS 5.0
```

```
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,467A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 2 009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 6063901 relevant
; ANTI-SENSE: No. 6063901 relevant
US-08-893-467A-1

Query Match          7.3%; Score 99.4; DB 3; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.8e-16;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGTCCACCGGGCCCGCCAGGAGGTCCAAAGAGGTGACAGAGATCCACAGGACCCCTGGC 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 721 CCAGGTGTCCAAGGACCCCGCCAGGCGCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 184 CCAACTGSCAACAAAGGACAAAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 781 CTTACGGGTGCACAGGGAACAGGAGCAACTGTGTCTCCAGGACCTCGAGGAGAGAAG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 244 GGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCGCCGAGAGCGTGGCGGCAAGGATCT 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 841 GGCAGCAAGGTGACATAGGTCTCAGTGGCCCAAGGGGGAACATGGCCACCAAGGGAGAC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 304 AAAGCTCCAGGGGCCCAAGAGTCCCGTGTTCCTCCCTGGGAAGCCCGCCCTCAGGGC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 901 AAAGGGGACCTAGGCTTCCAGGAAACAAAGGGGACATGGGCATGAAGGAGACACAGGGG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 364 CCCAGTGGGACCCAGGCGCCCGCCAGGACCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 961 CCCATGGGGTCCCTGGAGCTCAGGGAGGTAAAGGTATGTGTGAAACACCGGCTTACCA 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 424 GGCCTCTCTGGCTTCCAGGGACTTCAGGCGACCCGTTGGGGAGCCCTGGGGTGGCTGGACCT 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1021 GGTGGTGGTGGATCTCCAGGAGTCAAGGTGACCAAGGAAACCTGGAGTCAGGGTGT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 484 CGGGAGTCCAGGCTTCCTGGGGTACAGGCATGCCAGGCCCAAGGGCCCGCCCGCCG 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1081 CCAGGCCCTCAAGGTGACAGGACTTTTCAGGTGCCAAGGGTGGAGCCAGGAGGAGGAGGAG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Qy 544 CTCTCTGGCCCATCAGGAGCGGTGG 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1141 CTTCTGGGCCAGGAGACCCCGG 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 13
US-08-555-669-11
; Sequence 11, Application US/08555669
; Patent No. 5773248
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
```


Search completed: March 21, 2003, 08:26:12
Job time : 66.7621 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 206.471 Seconds
(without alignments)
5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_2024

Perfect score: 1355

Sequence: 1 atgcaacaagattgatgag.....agaaaaaggttatcatcccg 1355

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1335.8	98.6	2930	10	US-09-745-763-198
c	249	18.4	415	10	US-09-954-456-1975
3	111	8.2	6728	10	US-09-954-456-782
4	111	8.2	6728	10	US-09-880-107-3946
5	110.6	8.2	1133	9	US-09-924-340-57
6	110.6	8.2	1133	9	US-09-952-600A-57
7	109.4	8.1	2691	10	US-09-925-302-64
8	107.2	7.9	2192	10	US-09-925-301-42
9	107	7.9	5086	10	US-09-880-107-3947
10	107	7.9	5145	9	US-09-925-299-206
11	107	7.9	5145	10	US-09-925-299-206
12	107	7.9	5432	12	US-10-044-090-22
13	106.8	7.9	4908	9	US-10-001-887-33
14	106	7.8	2542	9	US-09-954-531-961
15	106	7.8	2542	10	US-09-964-824A-255
16	105.4	7.8	5416	10	US-09-954-456-786
17	105.4	7.8	5416	10	US-09-880-107-2094
18	105.2	7.8	6158	10	US-09-919-497-6
19	105.2	7.8	6158	10	US-09-954-456-762

20	104.2	7.7	4149	12	US-10-044-090-104	Sequence 104, App
21	102.4	7.6	1486	10	US-09-925-302-247	Sequence 247, App
22	101.6	7.5	3226	10	US-09-954-456-725	Sequence 725, App
23	101.4	7.5	3690	12	US-10-044-090-448	Sequence 448, App
24	98.6	7.3	2158	9	US-10-001-835-98	Sequence 98, Appl
25	97.8	7.2	2520	10	US-09-880-107-3685	Sequence 3685, Ap
26	95.8	7.1	1485	10	US-09-925-302-246	Sequence 246, App
27	95.2	7.0	1797	9	US-09-978-295A-613	Sequence 613, App
28	95.2	7.0	1797	9	US-09-978-697-613	Sequence 613, App
29	95.2	7.0	1797	9	US-09-978-192A-613	Sequence 613, App
30	95.2	7.0	1797	9	US-09-999-832A-613	Sequence 613, App
31	95.2	7.0	1797	9	US-09-978-189-613	Sequence 613, App
32	95.2	7.0	1797	9	US-10-174-590-331	Sequence 331, App
33	95.2	7.0	1797	9	US-10-176-758-331	Sequence 331, App
34	95.2	7.0	1797	9	US-10-175-737-331	Sequence 331, App
35	95.2	7.0	1797	9	US-10-173-706-331	Sequence 331, App
36	95.2	7.0	1797	9	US-10-175-738-331	Sequence 331, App
37	95.2	7.0	1797	9	US-10-175-752-331	Sequence 331, App
38	95.2	7.0	1797	9	US-10-176-482-331	Sequence 331, App
39	95.2	7.0	1797	9	US-10-176-757-331	Sequence 331, App
40	95.2	7.0	1797	9	US-10-176-913-331	Sequence 331, App
41	95.2	7.0	1797	9	US-10-180-552-331	Sequence 331, App
42	95.2	7.0	1797	9	US-10-180-557-331	Sequence 331, App
43	95.2	7.0	1797	9	US-10-173-700-331	Sequence 331, App
44	95.2	7.0	1797	9	US-10-174-572-331	Sequence 331, App
45	95.2	7.0	1797	9	US-10-174-579-331	Sequence 331, App

ALIGNMENTS

RESULT 1

US-09-745-763-198
; Sequence 198, Application US/09745763
; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
McCoy, John M.
LaVallie, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Treaacy, Maurice
Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 198:

SEQUENCE CHARACTERISTICS:

LENGTH: 2930 base pairs

TYPE: nucleic acid

; STRANDEDNESS: double									
; TOPOLOGY: linear									
; MOLECULE TYPE: cDNA									
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:									
US-09-745-763-198									
Query Match									
Best Local Similarity 98.6%; Score 1335.8; DB 10; Length 2930;									
Matches 1348; Conservative 0; Mismatches 2; Indels 1; Gaps 1;									
QY	1	ATGCAACAGATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT	60						
DB	1266	ATGCAACAAGATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT	1325						
QY	61	ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATA	120						
DB	1326	ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATA	1385						
QY	121	CTAAGAGTCCACCGGGCCCCAGGGTCCAAGAGTGACAGAGGATCCAGGAGCCCCCT	180						
DB	1386	CTAAGAGTCCACCGGGCCCCAGGGTCCAAGAGTGACAGAGGATCCAGGAGCCCCCT	1445						
QY	181	GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGAGGGGAGCCTGGACCACTGGCCCT	240						
DB	1446	GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGGGGAGCCTGGACCACTGGCCCT	1505						
QY	241	GGGGTGTAGAGAGGCCCCAATTTGGACAGCTGGTCCCGGAGAGCGTGGCGCAAGGA	300						
DB	1506	GGGGTGTAGAGAGGCCCCAATTTGGACAGCTGGTCCCGGAGAGCGTGGCGCAAGGA	1565						
QY	301	TCTAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG	360						
DB	1566	TCTAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG	1625						
QY	361	GGCCCCAGTGGGACCCAGGCCCCCGGCCACCAGGCAAGAGGAGCTCCCGGCCCT	420						
DB	1626	GGCCCCAGTGGGACCCAGGCCCCCGGCCACCAGGCAAGAGGAGCTCCCGGCCCT	1685						
QY	421	CAGGCCCCCTCTGGCTTCCAGGAGCTTCCAGGACACCGTTGGGGAGCCTGGGTGCTGA	480						
DB	1686	CAGGCCCCCTCTGGCTTCCAGGAGCTTCCAGGACACCGTTGGGGAGCCTGGGTGCTGA	1745						
QY	481	CTCGGGGACTCCAGGCTTGGCTGGGTACCAGGATGCCAGGCCCCAAGGGCCCCCCC	540						
DB	1746	CCTCGGGGACTCCAGGCTTGGCTGGGTACCAGGATGCCAGGCCCCAAGGGCCCCCCC	1805						
QY	541	GGCCCTCTCTGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAAGAGCAACCCCG	600						
DB	1806	GGCCCTCTCTGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAAGAGCAACCCCG	1865						
QY	601	GCACGGAGGACAATGGCTGCGCCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT	660						
DB	1866	GCACGGAGGACAATAGCTGCGCCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT	1925						
QY	661	TTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAGACAAGCTTCA	720						
DB	1926	TTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAGACAAGCTTCA	1985						
QY	721	CATCTTGTGTTTTCATAAAGACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGG	780						
DB	1986	CATCTTGTGTTTTCATAAAGACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGG	2045						
QY	781	AGAGAGAGCCATGGATCGGCTCAGACTCAGAGCTGAGAGCTGAAATGAATGGAAGTGGGTG	840						
DB	2046	AGAGAGAGCCATGGATCGGCTCAGACTCAGAGCTGAGAGCTGAAATGAATGGAAGTGGGTG	2105						
QY	841	GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCAT	900						
DB	2106	GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCAT	2165						
QY	901	GGCCATGGCCAGGAGAGACTGTCTGGTTGATTATGCTGGCAGTGGAAACCATTTTC	960						
DB	2166	GGCCATGGCCAGGAGAGACTGTCTGGTTGATTATGCTGGCAGTGGAAACCATTTTC	2225						

QY	961	CAATGTGAAGACGTCAATAAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTCT	1020
DB	2226	CAATGTGAAGACGTCAATAAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTCT	2285
QY	1021	GCATTATAACGAGCTGTGATGGATACATGAGCAAAATTTTCAGCTCTCAAGGCAAGG	1080
DB	2286	GCATTATAACGAGCTGTGATGGATACATGAGCAAAATTTTCAGCTCTCAAGGCAAGG	2345
QY	1081	ACACTCCCTTTCTAATTTGCATCACTTCTCATCAGATTGAAAAAAGAAAGCACTCAAAA	1140
DB	2346	ACACTCCCTTTCTAATTTGCATCACTTCTCATCAGATTGAAAAAAGAAAGCACTCAAAA	2404
QY	1141	CCAATTACTGAAAAAATTTGACAGTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC	1200
DB	2405	CCAATTACTGAAAAAATTTGACAGTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC	2464
QY	1201	TTGGGAACATAAATGTTCCCAAGGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA	1260
DB	2465	TTGGGAACATAAATGTTCCCAAGGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA	2524
QY	1261	TCACATAGATTCTCTCCGTCCGTACGTAACCGTGCATTATACAAATATGCTTCCAAAGTA	1320
DB	2525	TCACATAGATTCTCTCCGTCCGTACGTAACCGTGCATTATACAAATATGCTTCCAAAGTA	2584
QY	1321	TGGAACACTCCCAATCAGAAAAAGGTTATCAT	1351
DB	2585	TGGAACACTCCCAATCAGAAAAAGGTTATCAT	2615

RESULT 2

US-09-954-456-1975/c
; Sequence 1975, Application US/09954456
; Patent No. US2002115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE OF INVENTION: Sets
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1975
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: n=a,t,g or c
US-09-954-456-1975

Query Match 18.4%; Score 249; DB 10; Length 415;

	Best Local Similarity	95.3%; Matches 266;	Pred. No. 1.3e-55; Conservative 0;	Mismatches 12;	Indels 1;	Gaps 1;
QY	1073	GGCAAAGGACACTCCTTTTCTAATTTGCATCACTTCCTCATCAGATTGAATAAAAAAAAGC	1132			
Db	415	GGCAAGGGACACNCTTCCTTAATGCATCACCTCCTCATCAGATTG-AAAAATAAAAGC	357			
QY	1133	ACTGAAACCACCAATTACTGTGAAAAAAATTGACAGTAGTGTGTTTTTACCATCCGTCATTAC	1192			
Db	356	ACTGAAACCACCAATTACTGTGAAAAAAATTGACAGTAGTGTGTTTTTACCATCCGTCATTNC	297			
QY	1193	CCAAAGACTTGGGAACATAAATGTGTTCCCAGGGTGATATGCTGATTTTCATTGTGCACAT	1252			
Db	296	CCAAAGACTTGGGAACATAAATGTGTTCCCAGGGTGATATGCTGATTTTCATTGTGCACAT	237			
QY	1253	GGACTGAAATCACATAGATTCTCCCTCGTCAGTACACCGTGGCATTTATACAATAATTATGCTTT	1312			
Db	236	GGACTGAAATCACATAGATTCTCCCTCGTCAGTACACCGTGGCATTTATACAATAATTATGCTTT	177			
QY	1313	CCAAAGTATGGAACACTCCAATCATGAAAAAGGTTTATCAT	1351			
Db	176	CCAAGTATGGAACACTCCAATCATGAAAAAGGTTTATCAT	138			

RESULT 3
US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Candi
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782

	Query Match	8.2%	Score 111;	DB 10;	Length 6728;
	Best Local Similarity	53.0%;	Pred. No. 9.1e-19;		
	Matches 237;	Conservative	0;	Mismatches 210;	Indels 0;
				Gaps	0;
QY	122	TACAAGGTCCACCGGGCCCGAGGGGTCCAAAGAGGTGACAGAGGATCCACAGGACCCCGCTG	181		
Db	721	TCCAAGGTCCCGCTGGTGAGCCTGGCGAGCCTGGAGCTTCAGGTCCCATGGTCCCCGAG	780		
QY	182	GCCCAACTGGCAACAAGGGGACAAAAGAGAGAGGGGGAGCCCTGGACCACTTGGCCCGCTG	241		

Db 781 GTCCCCCAGGTCCCCCTGGAAAGAATGGAGATGATGGGAAGCTGGAAACCTGGTCGTC 840

Qy 242 CGGGTGAGAGAGGCCAAATTTGGACCAAGCTGGTCCCCCGAGAGAGCGTGGCGCAAGGAT 301

Db 841 CTGCTGAGCGTGGCCCTCTCGGCCCTCAAGGGTCTCGAGGATGTCGCCGAACAGCTGGCC 900

Qy 302 CTAAGAGCTCCCAAGGCCCAAGAGCTCCCGCTGGTTCCCTGGGAAGCCCGGCCCTCAGG 361

Db 901 TCCCTGGAATCAGGGACACAGAGTTTCAGTGGTTGGATGGTCCCAAGGAGATGCTG 960

Qy 362 GCCCACTGGGGACCCAGGCCGCCCGGGCCCCACCAAGGCAAGAGGAGCTCCCGGCCCTC 421

Db 961 GTCCTGCTGGTCTAAGGCTGAGCCCTGGCAGCCCTGGTGAAATGAGCTCCTGCTCAGA 1020

Qy 422 AGGGCCCTCTCGGTTCCAGGAGACTTCAGGGACACCGTTGGGGAGCCTGGGGTGGCTGGAC 481

Db 1021 TGGGCCCCCTGGCCCTGCCTGGTGTAGAGAGGTGCGCCTGGAGCCCTGGCCCTGCTGGTG 1080

Qy 482 CTCGGGAGCTGCCAGGCTTGCCTGGGGTACCAGGAGATGCCAGGCCCCCAAGGGCCCCCCC 541

Db 1081 CTCGTGGAATGATGTGCTACTGTGCTGCGCGGGCCCCCTGGTCCCAACCGGCCCGCTG 1140

Qy 542 GCCCTCTCGGCCCATCAGGAGCGGTGG 568

Db 1141 GTCCTCTGGCTTCCCTGGTGTGCTGTTG 1167

RESULT 4

US-09-880-107-3946

; Sequence 3946, Application US/09880107

; Patent NO. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherif, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3946

; LENGTH: 6728

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274615

US-09-880-107-3946

	Query Match	8.2%	Score 111;	DB 10;	Length 6728;
	Best Local Similarity	53.0%;	Pred. No. 9.1e-19;		
Matches	237; Conservative	0;	Mismatches 210;	Indels	0; Gaps,
Qy	122	TACAAAGGTCCACCGGGCCCCCAGGGGTC	CACAAGAGGTGCACAGGAGATCC	CAGGAGACCCTTG	181
Db	721	TCCAAGGTCCCCCTGGTGAAGCTGCCGAGCCTGGAGCTTCAGGTCCCATTGGTCCCCGAG			780
Qy	182	GCCCCAATCGCAACAAAGGACAGAGAAGAGGAGGAGGAGCCTGGACACCATGSCCCTG			241
Db	781	GTCCCCCAGGTCCCCCTGGAAGAATGGAGATGATGGGGAAGCTGGAACACCTGGTTCGTC			840
Qy	242	CGGGTGACAGAGGCCCAATTGGACACAGCTGGTCCCCCGGAGAGCGTGGCGCAAAGGAT			301
Db	841	CTGGTGAGCGTGGGCTCTCTGGGCTCTCAGGGTCTCGAGGATTGCCCGGAACAGCTGCC			900
Qy	302	CTAAAGGCTCCACGGGCCCCAAAGGCTCCCCTGGTTCCCTCTGGGAAGCCCGGCCCTCAGG			361
Db	901	TCCCTTGAATGAAGGCACACAGGTTTCAGTGGTTTTGGATGGTCCCAAGGAGATGCTG			960


```

Db 818 GGGGACAGGACAGAGGATCAAGAGGTACCGTGGGTCTCTGGCTCCAGGGTCC- 876
QY 367 AGTGGGAGCCAGGCCCCCGGCCCCCAGGCAAGAGAGGACTCCCGCGCCCTCAGGGC 426
Db 877 -----CCCTGGCCCTCCTGGCTCTCTCTGGTGAACAGGTCCCTCTGGAGCCTCTGGT 928
QY 427 CCTCTGGGTTCACAGGACTTCAGGCGACCGTGTGGGAGGACCTGGGTGCTGGACCTCGG 486
Db 929 CCTGTGGTCCCGGAGGTCCCGTCTGTCTGTGGTCTCTGGCAAGATGGACTCAAC 988
QY 487 GGAGTCCAGGCTTGGCTGGGTACCGATGCCAGGCCCCCAAGGGCCCCCGGCCCT 546
Db 989 GGTCTCCCTGGGCCCCAFTGGGCCCCCTGTCTCTGGGTGCGCACTGGTGTATGCTGCTCT 1048
QY 547 CTGGGCCCATCAGGAGCGG-TGTTGCCCTCGCCCTGCGAGAAATGAGCCAAACCGGCAAC 605
Db 1049 GTTGGTCCCCCGGCCCTCTCTGGACCTCTGTGTCCTGGTCCCTGGTCTCCAGCGCTGTTTC 1108
QY 606 GGAGGACAATGGTGGCCCGCTCACTGGAAGA 637
Db 1109 GACTTCAGCTTCTCGCCCGCCAGCCACCTCAAGA 1140

RESULT 9
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Query Match 7.9%; Score 107; DB 10; Length 5086;
Best Local Similarity 51.3%; Pred. No. 8.7e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 120 ACTACAGGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCACGGGACCCCC 179
Db 709 ACAGCCCGTGTCTCTGGTGTGAAGGTGAACCTGGTGGCCCTGGTGAAGTGAAGTCC 768
QY 180 TGGCCCACTGCTCAACAAGGACAGAAAGAGAGAGAGAGGAGGAGCTGGACCACTGGCCC 239
Db 769 AGGTCAACAGGAGCGCGTGGCTCTCTGGTGAAGAGACCTGTTGGTGCCCTGGCCC 828
QY 240 TGGGGTGAAGAGGCCCCAATTTGGACCACTGGTCCCGCCCGGAGAGCGTGGCGGCAAGG 299
Db 829 AGCTGGTGGCGGCGAGTGAAGTGTGGGTGCGCGTGGGTCTCTGGTGGTCCCTATGG 888
QY 300 ATCTAAAGCTCCCGGGGCCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGCCCTCA 359
Db 889 GTCTGTGGCCCTCCAGGCTTCCAGGTGCCCTGGCCCCCAAGGGTGAATTTGGAGCTGT 948
```

```

QY 360 GGGCCCCCACTGGGGAACCCAGGCCCCCCCCGGCCCCACAGGCAAAAGAGGAGTCTCCCCGGCCC 419
Db 949 TGTAAACGCTGTGCTGCTGGTCCCGCGGTCCCGCGTCCCGTGGTGAAGTGGTCTTCCAGGCT 1008
QY 420 TCAGGGCCCTCCTGGCTTCCAGGACTTCAGGCAACCGTTCAGGAGCCCTGGGGTGGCTGG 479
Db 1009 CTCGGCCCCCGTTGGACCTCTCTGGTAATCTCTGGAGCAACGGCCTTACTGGTGCCAAGG 1068
QY 480 ACTCGGGGACTGCCAGGCTTGCTGGGTATCCAGCATGCCAGGCCCCCAAGGGCCCCCCC 539
Db 1069 TGTGCTGGCTTCCCGGGTGTCTGGGGCTCCCGGCTCCCTGGAGCCCGCGGTATTCC 1128
QY 540 CGGCCCCCTCTCGCCCATCAGGAGCGGTGTGTCCTGGCCCTTGCAGAGTGAAGCAACCCC 599
Db 1129 TGGCCCTGTGTGCTGCTGGCGGTCTACTGTCAGAGAGACTTGTGTGGTGGCTGGTCC 1188
QY 600 GGC 602
Db 1189 AGC 1191

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 7.9%; Score 107; DB 9; Length 5145;
Best Local Similarity 51.3%; Pred. No. 8.8e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 120 ACTACAGGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGACCCCC 179
Db 712 ACAGCCCGTGTCTCTGGTGTGAAGGTCAACCTGGTGGCTCCCTGGTGAAGTGAAGTGA 771
QY 180 TGGCCCACTGCTCAACAAGGACAGAAAGAGAGAGGAGGAGCTGGACCACTGGCCC 239
Db 772 AGGTCAACAGGAGCGCGTGGGCTTCTCTGGTGAAGAGACGCTGTGGTGCCCTGGCCC 831
QY 240 TCGGGTGAAGAGGCCCCAATTTGGACCACTGGTCCCGCCCGGAGAGCGTGGTGCGCCCTGGCCC 891
Db 832 AGCTGGTGGCCCTGGCAGTGAAGTGTGGGTCCCGTGGGTCTCTGCTGGTCCCATTTG 891
```



```
QY 542 GCCCTCCTGGCCCATCAGAGCGGTGGTGCCTCC 574
    ||| ||||| | ||||| | |||||
Db 1050 GTCCCCCTGGGAAGAGGAGGAGCCAGGCGCTC 1082

RESULT 15
US-09-964-824A-255
; Sequence 255, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrikan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 255
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-255

Query Match      7.8%; Score 106; DB 10; Length 2542;
Best Local Similarity 52.1%; Pred. No. 1.le-17;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 122 TACAAGGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCTG 181
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 630 TCCCTGGACCCCGGTCCCGAGGCATCAGGGGTACCCAGGCATGGCAGGGCCCCAAGG 689

QY 182 GCCCAACTGGCAACAGGACAGAAAGGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTG 241
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 690 GAGAGCGGGCCCTCATGGATATAAGGCATGCTGGGCGCTATCGTGCCTGGCCAC 749

QY 242 CGGGTGAGAGGCGCCATTTGGACAGCTGTTCCCCCGGAGAGCGTGGCGGCAAGGAT 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 750 CGGGTGAGGAGGTCTTAGGGACCGCCAGGCGGAGCTGGGGAAGGAGGTGACGAGGCA 809

QY 302 CTAAAGGCTCCAGGCGCCCAAGGCTCCCGTGTTCCTCTGGGAAGCCCGGCCCTCAGG 361
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 810 GCCCAGGTATTCTGTGACCCCGAGGGGATCACAGGCCCCGAAAGGAGCAACGGGCCCCCAG 869

QY 362 GCCCCAGTGGGACCCAGGCCCGCCCGGCCACCAAGGAGGAGTCCCGGCGCCCTC 421
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 870 GCATCAACGGCAAGCATGGGACCCAGGCACGCTTGGCATGAAGGCGAGTGCAGGACAGG 929

QY 422 AGGGCCCTCCTGGCTTCAGGAGCTTCAGGGCACCGTTGGGAGCGCTGGGTGCTGAC 481
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 930 CGGGACAGCCCGGAAGTCCAGGCCACCAAGGGGCTAGCGGCTGTCAGGCCAGCCTGGGA 989

QY 482 CTCGGGAGTCCAGGCTTGGCTGGGTACCAGGCATGCCAGGCCCCCAAGGCCCCCGG 541
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 990 CAAAAGGAGGCCCTGGACACCAAGGATGAGCCGGGCCCCAGGGCCCTTCTGGATTCTG 1049

QY 542 GCCCTCCTGGCCCATCAGAGCGGTGGTGCCTCC 574
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1050 GTCCCCCTGGGAAGAGGAGGAGCCAGGCGCTC 1082
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 1777.64 Seconds
(without alignments)
12344.923 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_2024

Perfect score: 1355

Sequence: 1 atgcacacagattgatgag.....agaaaaaggtatcatcccg 1355

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_Other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	900	66.4	3305	11	BC009162	BC009162 Mus muscu
c 2	762.4	56.3	967	9	AL568743	AL568743
3	643.8	47.5	668	14	BQ674807	BQ674807 AGENCOURT
4	556.4	41.1	861	14	BQ713873	BQ713873 AGENCOURT
c 5	523.4	38.6	723	14	BQ004176	BQ004176 UI-H-EIO-
6	514.6	38.0	552	10	AW958053	AW958053 EST370123

c	7	499.2	36.8	655	12	BG482931	BG482931 602502939
	8	469.8	34.7	500	14	BM676508	BM676508 UI-E-EJO-
	9	467.4	34.5	507	14	BM713891	BM713891 UI-E-EJO-
	10	462.6	34.1	906	13	BI456109	BI456109 603172765
	11	445.6	32.9	580	12	BE910803	BE910803 601661855
	12	441.8	32.6	808	14	BQ771366	BQ771366 UI-M-FIO-
c	13	416.8	30.7	517	9	AL543000	AL543000 AL543000
	14	415.8	30.8	601	10	BE290299	BE290299 601089246
	15	406	30.0	638	10	BB248064	BB248064 BB248064
	16	403.4	29.8	957	14	BQ955927	BQ955927 AGENCOURT
	17	353.4	26.1	504	14	R74387	R74387 yi57f11.r1
c	18	332.4	24.5	656	14	BQ021066	BQ021066 UI-H-DHI-
c	19	328.8	23.5	345	9	AI742661	AI742661 wg44f06.x
	20	318.4	23.5	352	10	AW956633	AW956633 EST368703
	21	318.4	23.5	552	10	AW956637	AW956637 EST368697
	22	314.6	23.2	421	14	R78202	R78202 yi81h07.r1
	23	314.2	23.2	455	10	AW023880	AW023880 df61c07.y
	24	311	23.0	339	9	AA304251	AA304251 EST17006
	25	308.2	22.7	943	14	BQ887163	BQ887163 AGENCOURT
	26	301.4	22.2	357	10	AW379436	AW379436 CMA-HT024
	27	297	21.9	936	14	BQ891432	BQ891432 AGENCOURT
c	28	296.4	21.9	334	10	AW002141	AW002141 wt84a12.x
	29	296	21.8	310	9	AA361740	AA361740 EST71089
c	30	289.6	21.4	318	12	BE767950	BE767950 QV1-GN006
	31	285.8	21.1	309	14	W76214	W76214 zd58609.r1
	32	271	20.0	350	9	AA012704	AA012704 RPU1402CG
	33	268.8	19.8	294	9	AA357879	AA357879 EST66937
c	34	259.6	19.2	288	14	W72977	W72977 zd58609.s1
	35	254.6	18.8	414	14	R27402	R27402 yH46c10.r1
c	36	252.6	18.6	498	9	AL569543	AL569543 AL569543
c	37	249	18.4	415	14	R74388	R74388 yi57f11.s1
	38	245.8	18.1	420	12	BE929813	BE929813 RC3-GN004
	39	236.2	17.4	441	10	BB749743	BB749743 BB749743
c	40	233.8	17.3	429	12	BE929798	BE929798 RC3-GN004
	41	229.4	16.9	608	10	AW918393	AW918393 EST349697
c	42	222.8	16.4	381	12	BE767917	BE767917 QV1-GN006
c	43	221	16.3	508	9	AA700864	AA700864 zj36b11.s
c	44	220.4	16.3	538	9	AI954661	AI954661 wg34f12.x
c	45	209	15.4	501	10	AW275816	AW275816 xq39c09.x

ALIGNMENTS

RESULT 1	BC009162	BC009162	3305 bp	mRNA	linear	HTC 07-AUG-2002
LOCUS	BC009162	Mus musculus, Similar to scavenger receptor with C-type lectin, clone IMAGE:2811487, mRNA.				
DEFINITION	BC009162	clone IMAGE:2811487, mRNA.				
ACCESSION	BC009162	BC009162.1	GI:14714370			
VERSION	BC009162	HTC.				
KEYWORDS	HTC.	house mouse.				
SOURCE	BC009162	Mus musculus				
ORGANISM	BC009162	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1	(bases 1 to 3305)				
AUTHORS		Strausberg, R.				
TITLE		Direct Submission				
JOURNAL		Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK		NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT		Contact: MGC help desk Email: cgpbbs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/				

Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 5 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: Incomplete processing.

FEATURES
source

Location/Qualifiers
1..3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 1031 a 757 c 760 g 757 t
ORIGIN

Query Match 66.4%; Score 900; DB 11; Length 3305;
Best Local Similarity 81.7%; Pred. No. 4.5e-173;
Matches 1106; Conservative 0; Mismatches 230; Indels 18; Gaps 5;

QY 1 ATGCACAAAGATTGATGAGGTGCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGAT 60
DB 1296 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCAGTGGTT 1355
QY 61 ATGGAGGAATCAAGCTAGTACAGTCCAGCATGCTCAGCTCATCAAGAAATTTTCAATA 120
DB 1356 ATGGAGAGATGAATGTTGTTCACTCCAAGCACGTTGCTCATCAAGAACTTTTACCAAT 1415
QY 121 CTACAGGTTCACCGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCAGGGACCCCT 180
DB 1416 CTACAGGTCTCTCTGGCCCCAGAGTCCAAGAGGTGACAGAGGATCTCAGGGACCACT 1475
QY 181 GGGCCACTGGCAACAAGGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTGGCCCT 240
DB 1476 GGTCCAACCTGGCAACAAGGGACAGAAAGGAGAGAGGGAGCCCTGGTCCACCTGGCCCT 1535
QY 241 GCGGGTGAGAGAGGCCCAATTCGACCAGCTGCTCCCGGAGAGCGTGGCGGCAAGGA 300
DB 1536 GCGGGTGAGAGGGGCACAAATTCGACGTCGGCCCTCTCTGGAGAGCGTGGCAGCAAGGA 1595
QY 301 TCTAAGGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 360
DB 1596 TCCAAAGGCTCACAGGTCCTCAAGAGATCTCTGGGTCCCCAGGGAAGCTTGGCCCTCAA 1655
QY 361 GCGCCCACTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGAGGACTTCCCGGCCCT 420
DB 1656 GGACCTAGTGGGGACCCAGGACCAACAGGTCCACCAAGGCAAGGATGGACTTCCCTGGCCCT 1715
QY 421 CAGGGCCCTCTCTGGCTTCAGGAGCTTCAGGCAAGCTTTCAGGAGCGCTGGGGTGGCTTGA 480
DB 1716 CAGGGCCCTCTCTGGCTTCAGGAGCTTACAGGCAAGCTTGGGGTGGAGTACTTGA 1775
QY 481 CCTCGGGGACTGCCAGGCTTGGCTGGGTACCAGGATGCCAGGCCCAAGGGCCCCCCC 540
DB 1776 CCTCGGGGTTCAGGCTTGGCAGGGTGGCCAGGATGCCCTGGGCTTAAGGGACCACT 1835
QY 541 GCGCCCTCTGGCCCATCAGGAGGGTGGTGGCCCTGGGCCCTGCAGATGAGCCAAACCCGG 600
DB 1836 GCGCCCTTCCAGGCCCCCTCAGGAGCAATGGAGCAATGGGCTCTGCAGAAATGAACCAACCCA 1895
QY 601 GCACCGGAGCAATGGCTGGCCGCTCTACTGAGAGACTTCACAGACAAATGCTACTAT 660
DB 1896 GCATCAGAGGTCAACGGATGTCCACCTCTACTGGAAGACTTCACAGATAATGCTACTAT 1955

QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCA 720
DB 1956 TTTTCAGTTGAGAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAGTCTTCC 2015
QY 721 CATCTTGTTCATTAACACTAGAGAGGAGCAACATGATGATAAAGACAGATGTTAGGG 780
DB 2016 CATCTCGTTTTCATTAACACTAAGAGAAACACATGATGATAAAGACATACCTGGGG 2075
QY 781 AGAGAGAGCACTGGATCGGCTCACAGACTCAGACGTCAGAAATGAATGAAGTGGCTG 840
DB 2076 AGAGAAAGCCATGGATCGGCTCACAGACTCAGAAAGGAGCAATGGAAGTGGCTA 2135
QY 841 GATGGACATCTCCAGACTTACAAAATTTGAAAGCTGGACAGCCGAGTAACCTGGGTCT 900
DB 2136 GACGGTCACTGTTGATTACAAAACCTGGAAGCTGGACACCAAGATAACTGGGCACT 2195
QY 901 GCCATGGCCAGGAGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGAAACGATTC 960
DB 2196 GCGCATGGCCAGGAGAGACTGTGCTGGCTTGTATTTAGCGAGGACAGTGAATGACTTC 2255
QY 961 CAATGTGAAGAGCTCAATAAATTTTCGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGG 1020
DB 2256 CAGTGTGATGAATCAATAAATTTTGTGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 2315
QY 1021 GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCA--GCTCTCAAAGGCCAA 1078
DB 2316 ATATTATAGCAGCATGATATAATAGCAGAAACATATTTTCTGATGCTCTGAAAGCCGAA 2375
QY 1079 GGACACTCTTCTTAATTCATCCTCTCTATCAGATGAAAAAAGGAGGAGGAGGAGGAGGAGG 1138
DB 2376 GAATGCTGTTCTGATTCATCA--CTTCTCACCAGATTG--AATGGAAAAAGCTCTGAA 2432
QY 1139 AACCATACTCAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCGTCAATACCCAAAG 1198
DB 2433 AGTAGTTATTTCAAAATTAATGGACACCTACTG-----CACAATACCCAAG 2480
QY 1199 ACTTGGGAACAAAATG--TTCCCGAGGTGATGATGCTGATTTTTCATTTGTCACATGGACT 1257
DB 2481 ACTAGGGGCTAAATGCTACCCCAAGTGTATATATTGATTTTCCAGTGTACAAATGGACT 2540
QY 1258 GAATCAGATAGATTCCTCCCTCAGTACCGTGGATATACAAATATATGCTTTCCAAA 1317
DB 2541 GAATGCGATAGATTTTCTCAGCCATTAACCATAGAATTTATGCAAAATATATCTTTCCAA 2600
QY 1318 GTATGGAACACTCCCAATCAGAAAAAGTTATCAT 1351
DB 2601 ATATGGAATGCTCCCAATCAGAAAAAGACTATCAT 2634

RESULT 2

AL568743/c

LOCUS

DEFINITION

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

AL568743

Db	241	CTAGAGGACAGCAATGGATATAAAAAACACATGGTAGGGAGAGAGCCACTGGATCG	300
QY	800	GCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACT	859
Db	301	GCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACT	360
QY	860	ACAAAAATTTGGAAGCTGGACAGCGGATAACTGGGGTCATGGCCATGGCCGAGGAGAAG	919
Db	361	ACAAAAATTTGGAAGCTGGACAGCGGATAACTGGGGTCATGGCCATGGCCGAGGAGAAG	420
QY	920	ACTGTGCTGGTGTGATTTATGCTGGGACGTGAACGATTTTCAATGTGAAGACGTCAATA	979
Db	421	ACTGTGCTGGTGTGATTTATGCTGGGACGTGAACGATTTTCAATGTGAAGACGTCAATA	480
QY	980	ACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATATTAACGGACTGTGA	1039
Db	481	ACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATATTAACGGACTGTGA	540
QY	1040	TGGGATCACATGAGCAATTTTCAGCTCTCAAGGCAAGGACATCTCTTCTTAATTGCA	1099
Db	541	TGGGATCACATGAGCAATTTTCAGCTCTCAAGGCAAGGACATCTCTTCTTAATTGCA	600
QY	1100	TCACCTTCTCATCAGATTCAAAAAAAGCACTGAAACCAATTTACTGAAAAAAA	1158
Db	601	TCACCTTCTCATCAGATTG-AAAAAAGCACTGAAACCAATTTACTGAAAAAAA	658
RESULT 4			
LOCUS	B0713873	861 bp	mRNA
DEFINITION	AGENCOURT_8492828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240	linear	EST 16-JUL-2002
ACCESSION	B0713873		
VERSION	B0713873.1	GI:21852772	
KEYWORDS	EST		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	NIH-MGC http://mgi.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: Susan L. Sullivan, PhD.		
	cDNA Library Preparation: ResGen, Invitrogen Corp		
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLAM13721 row: n column: 01		
	High quality sequence stop: 598.		
FEATURES	Location/Qualifiers		
source	1. .861		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:6306240"		
	/clone_lib="NIH_MGC_129"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: olfactory epithelium; Vector:		
	pCMV-SPORT6.1.ccdB; Site_1: EcorV; Site_2: NotI; Cloned		
	unidirectionally. Primer: Oligo dt. Average insert size		
	2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this		
	is a NIH_MGC Library."		
BASE COUNT	221 a 253 c 239 g 147 t	1	others
ORIGIN			
Query Match	41.1%	Score 556.4;	DB 14;
Best Local Similarity	82.8%	Pred. No. 3.7e-103;	Length 861;
Matches 647;	Conservative 0;	Mismatches 132;	Indels 2;
			Gaps 1;

QY	14	TGATGAGGTGCGAGGTAGACACTGAAGTAGCCAACTTTATCACTGATTTATGGAGAANAATGA	73
Db	14	TGATGAGGTCAAAGTTAGACACTGAAGTGCCCAACTTATCACTGTTATGGGAAGATGA	73
QY	74	AGCTAGTAGACTCCAAAGCATGCTCAGCTCAACAAGATTTTACAATACTACAAAGGTCCAC	133
Db	74	AACTGGTTGACTCCAAGCACGCTCAGCTCATCAAGAATTTTACCATTCTACAAGGTCCCTC	133
QY	134	CGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGGACCCCTGGCCCCAACTGGCA	193
Db	134	CTGGCCCCAGAGTCCAAAAGGTGACAGAGATCTCAGGGACCACTGGTCCAACTGGCA	193
QY	194	ACAAGGACAGAAAGAGAGAGAGGGAGCCTGGACCACTGGCCCTGGGGTGAGAGAG	253
Db	194	ACAAGGACAGAAAGAGAGAGAGGGAGCCTGGTCCACCTGGCCCTGGGGTGAGAGG	253
QY	254	GCCCAATTTGGACCAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAGGCTCCC	313
Db	254	GCACAATTTGGACCACTCGGCCCTCTGGAGAGCGTGGCAGCAAGGATCCAAAGGCTCAC	313
QY	314	AGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGG	373
Db	314	AGGGTCCCAAGAGATCTCGTGGTCCCCAGGAGAGCCTGGCCCTCAAGGACCTAGTGGGG	373
QY	374	ACCCAGGCCCCCGGGCCACACAGGCAAGAGGGACTTCCCCGGCCCTCAGGGCCCTCTCG	433
Db	374	ACCCAGGACCACAGTCCACCAGCAAGGATGGACTCCCTGGCCCTCAGGGCCCTCTCG	433
QY	434	GCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCTTGGGGTGGCTGGACCTCGGGGACTGC	493
Db	434	GCTTCCAGGGACTACAGGGCACTTGGGTGAGCCTGGAGTACCTGGACCTCGGGGGTTGC	493
QY	494	CAGGCTTGGCTGGGTACCAAGCATGCCAGGCCCCAAGGGCCCCCGGCCCTCTCTGGCC	553
Db	494	CAGGCTTGGCAGGGGTGCCAGGCATGCCCTGGGCCCTAAGGGACCACTGGCCCTCCAGGC	553
QY	554	CATCAGGAGCGGTGGTGGCCCTGCGCAATAGAGCAACACCCGGACCGGAGGACA	613
Db	554	CCTCAGGAGCAATGGAGCCATTGGCTCTGCAAAATGAACCAACCCCAATCAGAGGTCA	613
QY	614	ATGGCTGCCCGCTCAGTGGGAAGACTTCACAGACAAATGCTACTATTTTTCAGTTGACA	673
Db	614	ACGGATGTCGGCTCAGTGGGAAGAACTTCACAGATAAATGCTACTATTTTCTTGGCA	673
QY	674	AAGAAATTTTGGAGATGCAAGC--TTTCTGTGGAAGACAAGTCTCTCACTCTTGT	731
Db	674	AAAAAATTTTCGAAAATGCTACCCCTTTTCTGCGAACAATAATCTCCCACTCCGGTTT	733
QY	732	CATAAAGCTAGAGAGACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCA	791
Db	734	CATAACCCCAAGAAAAAACGCCATGGGATAAAAAACCTACCTTTCGGGGACCAAGAAC	793
QY	792	C 792	
Db	794	C 794	
RESULT 5			
LOCUS	B0004176/c	723 bp	mRNA
DEFINITION	UI-H-EIO-aym-p-04-0-UI.sl NCI-CGAP_EIO Homo sapiens cDNA clone	linear	EST 26-MAR-2002
ACCESSION	B0004176		
VERSION	B0004176.1	GI:19729076	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 723)		
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-ri@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1. 723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5840883"
/clone_lib="NCL_CGAP_E10"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
with a modified polylinker; Site:1. EcoR I; Site:2. Not I;
NCL_CGAP_E10 is a cDNA library containing the following
tissues): Chondrosarcoma. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is ACACCTGCAC.
TAG_LIB=UI-H-E10
TAG_TISSUE=chondrosarcoma
TAG_SEQ=ACACTTGCAC"

BASE COUNT 185 a 148 c 130 g 260 t
ORIGIN

Query Match 38.6%; Score 523.4; DB 14; Length 723;
Best Local Similarity 99.5%; Pred. No. 1.9e-96;
Matches 546; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
QY 803 TCACAGACTCAGAGCGTGAATAATGAATGGTGGTGGATGGGACATCTCCAGACTACA 862
DB 723 TCACAGACTCAGAGCGTGAATAATGAATGGTGGTGGAT-GGACATCTCCAGACTACA 665
QY 863 AAAATTGGAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGGCCAGGAGAGACT 922
DB 664 AAAATTGGAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGGCCAGGAGAGACT 605
QY 923 GTGCTGGTGTATTTATGCTGGCAGTGAACGATTTCCAATGTGAAGACGTCAATAACT 982
DB 604 GTGCTGGTGTATTTATGCTGGCAGTGAACGATTTCCAATGTGAAGACGTCAATAACT 545
QY 983 TCATTTCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTATAACGGACTGTGATGG 1042
DB 544 TCATTTCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTATAACGGACTGTGATGG 485
QY 1043 GATCATATGAGCAAAATTTTACGCTCTCAAAGGCAAGGACACTCCCTTTCTAATTCATCA 1102
DB 484 GATCATATGAGCAAAATTTTACGCTCTCAAAGGCAAGGACACTCCCTTTCTAATTCATCA 425
QY 1103 CCTTCTCATCATGTAATAAAAAAGACACTGAAACCAATTAAGTGAATAAATTTGA 1162
DB 424 CCTTCTCATCATGTTG-AAAAAAGAAAGACACTGAAAGCCAAATTAAGTGAATAAATTTGA 366
QY 1163 CAGCTAGTGTCTTTTACCATCCGTCATTACCAAGACACTTGGGAACATAAATGTTCCCA 1222
DB 365 CAGCTAGTGTCTTTTACCATCCGTCATTACCAAGACACTTGGGAACATAAATGTTCCCA 306
QY 1223 GGGTGATATGCTGATTTTCAATTTGTGCACATGGACTGAATCAGATATCTCCTCCGTCA 1282

DB 305 GGGTGATATGCTGATTTTTCATTTGTCACATGGACTGAATCAGATAGATTCTCCTCCGTCA 246
QY 1283 GTAACCGTCGATTTATACAAATTTATGCTTTCAAAAGTATGGAACTCTCAATCAGAAAAA 1342
DB 245 GTAACCGTCGATTTATACAAATTTATGCTTTCAAAAGTATGGAACTCTCAATCAGAAAAA 186
QY 1343 GCTTATCAT 1351
DB 185 GCTTATCAT 177

RESULT 6
AW958053
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
AW958053
EST370123
MAGE resequenes, MAGE Homo sapiens cDNA, mRNA sequence.
AW958053
AW958053.1
GI:8147736
EST.
SOURCE

REFERENCE
AUTHORS
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL
COMMENT
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 115
Seq primer: Reverse.

FEATURES
source
Location/Qualifiers
1. 552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequenes, MAGE"
/note="Vector: pBluescriptSKm"
BASE COUNT 176 a 112 c 143 g 121 t
ORIGIN

Query Match 38.0%; Score 514.6; DB 10; Length 552;
Best Local Similarity 98.0%; Pred. No. 1.2e-94;
Matches 542; Conservative 0; Mismatches 9; Indels 2; Gaps 2;
QY 574 CTGGCCCTCAGAAATGAGCAACCCCGCCAGGAGCAATGGCTGCCCTCACTGG 633
DB 1 CTGGCCCTCAGAAATGAGCAACCCCGCCAGGAGCAATGGCTGCCCTCACTGG 60
QY 634 AAGAATCTCACAGACAAATGCTACTATTTTTCAGTTTGAGAAAGAAATTTTGAGGATGCA 693
DB 61 AAGAATCTCACAGACAAATGCTACTATTTTTCAGTTTGAGAAAGAAATTTTGAGGATGCA 120
QY 694 AGCTTTTCTGTGAGACAAAGTCTTCACATCTTGTGTTTTCATAAACAAGTGTGAGGAGGACAG 753
DB 121 AAGCTTTTCTGTGAGACAAAGTCTTCACATCTTGTGTTTTCATAAACAAGTGTGAGGAGGACAG 180
QY 754 CAATGGATAAAAAAGACAGATGGTAGGAGAGAGAGCACTGGATCGCCTCACAGACTCA 813
DB 181 CAATGGATAAAAAAGACAGATGGTAGGAGAGAGAGCACTGGATCGCCTCACAGACTCA 240
QY 814 GAGCGTGAATAATGAAATGGCTGGATGGGACACTCTCCAGACTACAAAAATTTGAAA 873
DB 241 GAGCGTGAATAATGAAATGGCTGGATGGGACACTCTCCAGACTACAAAAATTTGAAA 300
QY 874 GTGGACAGCCGGATAACTGGGGTCAATGGCCATATGGCCAGGAGAGAGACTGTGCTGGGTG 933
DB 301 GCTGGACAGCCGGATAACTGGGGTCAATGGCCATATGGCCAGGAGAGAGACTGTGCTGGGTG 360

```
QY 934 ATTTATGCTGGCAGTGGAGCAATTTCCAAATGTGAAGAGCTCAATAAC-TTCATTTTGGCA 992
|||||
Db 361 ATTTATGCTGGCAGGGAAGCAATTTCCAAATGTGAAGAGCTCAATAAC-TTAATTTGGCA 420
|||||
QY 993 AAAGACAGGGAGACAGTACTGTCATCTGCATTTATTAACGGAGCTGGATGCATGACATGA 1052
|||||
Db 421 AAAGACAGGGAGACAGTACTGTCATCTGCATTTATTAACGGAGCTGGATGCATGACATGA 480
|||||
QY 1053 GCAAAATTTTCAGCTCTCAAGGCAAGGACACTCCCTTTCTTAATTTGCATCACCTTTCATC 1112
|||||
Db 481 GCAAAATTTTCAGCTCTCAAGGCAAGGAC-CTCCCTTTCTAATTTGGCTCACCTTTCATC 539
|||||
QY 1113 AGATTGAAAAAAA 1125
|||||
Db 540 AGATGAAAAAAA 552
|||||

RESULT 7
BG482931
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BG482931 655 bp mRNA linear EST 21-MAR-2001
602502939F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4616625 5',
mRNA sequence.
BG482931
BG482931.1 GI:13415210
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 655)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-re@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA library Preparation: CLONTECH Laboratories, Inc.
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1371 row: m column: 10
High quality sequence stop: 651.
Location/Qualifiers
1. .655
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4616625"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
sfii (ggccgctcgcc); Site:2: sfii (ggccattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 219 a 124 c 144 g 167 t 1 others
ORIGIN

Query Match 36.8%; Score 499.2; DB 12; Length 655;
Best Local Similarity 98.1%; Pred. No. 1.6e-91;
Matches 516; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 826 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAACCTGGACAGCCG 885
|||||
Db 1 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAACCTGGACAGCCG 60
|||||
```

```
QY 886 GATAACTGGGGTCATGGCCATGGCCAGGAGAGACTGCTGGGGTTGATTTATGCTGGG 945
|||||
Db 61 GATAACTGGGGTCATGGCCATGGCCAGGAGAGACTGCTGGGGTTGATTTATGCTGGG 120
|||||
QY 946 CAGTGAAGCAATTTCAATGTGAAGAGCTCAATAACTTTCATTTGCCAAAAAGACAGGAG 1005
|||||
Db 121 CAGTGAAGCAATTTCAATGTGAAGAGCTCAATAACTTTCATTTGCCAAAAAGACAGGAG 180
|||||
QY 1006 ACAGTACTGTCATCTGCATTTAATTAACGGAGCTGGATGGGATGCATGAGCAAAATTTTCAGC 1065
|||||
Db 181 ACAGTACTGTCATCTGCATTTAATTAACGGAGCTGGATGGGATGCATGAGCAAAATTTTCAGC 240
|||||
QY 1066 TCTCAAAGGCAAGGACACTCCTTTCTTAATTTGCATCACCTTCTCATCAGATTGAAAAAAA 1125
|||||
Db 241 TCTCAAAGGCAAGGACACTCCTTTCTTAATTTGCATCACCTTCTCATCAGATTG--AAAAA 298
|||||
QY 1126 AAAAAAGCACTGAAAAACCAATTACTGAAAAAAATTTACAGAGCTAGTCTTTTACCATCCG 1185
|||||
Db 299 AAAAAAGCACTGAAAAACCAATTACTGAAAAAAATTTACAGAGCTAGTCTTTTACCATCCG 358
|||||
QY 1186 TCATTACCCCAAGACTTGGGAACCTAAATGTTCCCGAGGGTGATATGCTGATTTTCATTG 1245
|||||
Db 359 TCATTACCCCAAGACTTGGGAACCTAAATGTTCCCGAGGGTGATATGCTGATTTTCATTG 418
|||||
QY 1246 TGCACATGGAGTGAATCACAATAGATTCTCTCGTCAGTAACCGTCGATATACAAAT 1305
|||||
Db 419 TGCACATGGAGTGAATCACAATAGATTCTCTCGTCAGTAACCGTCGATATACAAAT 478
|||||
QY 1306 ATGTCTTCCAAAGTAGTGAACACTCCCAATCAGAAAAAGGTTATCAT 1351
|||||
Db 479 ATGTCTTCCAAAGTAGTGAACACTCCCAATCAGAAAAAGGTTATCAT 524
|||||

RESULT 8
BM676508/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BM676508 500 bp mRNA linear EST 27-FEB-2002
UI-E-EJ0-ahq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone
UI-E-EJ0-ahq-c-05-0-UI 3', mRNA sequence.
BM676508
BM676508.1 GI:18986404
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 500)
Bonald, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA=Yes
Location/Qualifiers
1. .500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="UI-E-EJ0-ahq-c-05-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
```

optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"

```

/dev_stage="fetal and adult"
/lab_host="PH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker: Site.1: EcoR I; Site.2: Not I;
UI-E-E10 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGCAATCAAGA
; lens, CGATTAGGCA; eye anterior segment, AATGCCGATG;
optic nerve, CCAATTAAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).
TAG LIB=UI-E-E10
TAG_TISSUE=human_eye
TAG_SEQ=AACTCCCAAT"
```

BASE COUNT	109 a	123 c	89 g	179 t	
ORIGIN					
Query Match	34.7%	Score 469.8	DB 14	Length 500	
Best Local Similarity	99.4%	Pred. No. 1.5e-85			
Matches 482	Conservative 0	Mismatches 2	Indels 1	Gaps 1	
QY	674	AAGAAATTTTGGAGTGCAAAAGCTTTCTGTGGAAGACAAGTCTTCACATCTTGTGTTTCA	733		
Db	494	AGGAAATTTTGGAGTGCAAAAGCTTTCTGTGGAAGACAAGTCTTCACATCTTGTGTTTCA	435		
QY	734	TAAACACTAGAGAGGAACACGATGGATAAAAAACAGATGGTAGGGAGAGAGAGCCACT	793		
Db	434	TAAACACTAGAGAGGAACACGATGGATAAAAAACAGATGGTAGGGAGAGAGAGCCACT	375		
QY	794	GGATCGGCCCTCACAGACTCAGAGCGTGCAAAATGAATGGAAGTGGCTGGATGGGACATCTC	853		
Db	374	GGATCGGCCCTCACAGACTCAGAGCGTGCAAAATGAATGGAAGTGGCTGGATGGGACATCTC	315		
QY	854	CAGACTACAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGGGCCAG	913		
Db	314	CAGACTACAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGGGCCAG	255		
QY	914	GAGAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGGGAACGATTTCCAATGTGGAAGACG	973		
Db	254	GAGAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGGGAACGATTTCCAATGTGGAAGACG	195		
QY	974	TCAATAACTTTCATTTGGGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTATAACCGA	1033		
Db	194	TCAATAACTTTCATTTGGGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTATAACCGA	135		
QY	1034	CTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAGCCAAAGGACACTCCTTTCTA	1093		
Db	134	CTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAGCCAAAGGACACTCCTTTCTA	75		
QY	1094	ATTGCATCACCTTCTCATCAGATTGAAAAAAGCACTGAAAAACCAATTACTGAAA	1153		
Db	74	ATTGCATCACCTTCTCATCAGATTGAAAAAAGCACTGAAAAACCAATTACTGAAA	16		
QY	1154	AAAAA 1158			
Db	15	AAAAA 11			
RESULT 9					
LOCUS	BM713891				
DEFINITION	UI-E-EJO-abq-c-05-0-UI.r1	507 bp	mRNA	linear	EST 28-FEB-2002
	UI-E-EJO-abq-c-05-0-UI 5', mRNA sequence.				

ACCESSION	BM713891
VERSION	BM713891.1
KEYWORDS	GI:19027149
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 507) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996)
JOURNAL	97044477
MEDLINE	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu
COMMENT	Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). The following repetitive elements were found in this cDNA sequence: 443-498, >POLY_A\$imple_repeat Seq primer: M13 Reverse.
FEATURES	Location/Qualifiers 1..507 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="UI-E-EJ0-abq-c-05-0-UI" /clone_lib="UI-E-EJ0" /tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid" /dev_stage="fetal and adult" /lab_host="DH10B (Life Technologies) (T1 phage resistant)" /note="Organ: eye; Vector: pW73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor. digested with Not I, and cloned directionally into pW73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, ACAATCAAGA ; lens, CGATTAGCGA; eye anterior segment, AATCCGCAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT	184 a 89 c 121 g 112 t 1 others
ORIGIN	
Query Match	34.5%; Score 467.4; DB 14; Length 507;
Best Local Similarity	99.4%; Pred. No. 4.7e-85;
Matches 479; Conservative	0; Mismatches 2; Indels 1; Gaps 1;
QY	677 AAATTTTGGAGTCGAAAGCTTTTCTGTGAAGACAAGTCTTTCATCTTGTTCATAA 736
Dd	1 AAAATTTTGGAGTCGAAAGCTTTTCTGTGAAGACAAGTCTTTCATCTTGTTCATAA 60
QY	737 ACAC TAGAGG AACCGCAATGGATATAAAAAACAGATGGTAGGAGAGAGGCCACTGGA 796
Dd	61 ACAC TAGAGG AACCGCAATGGATATAAAAAACAGATGGTAGGAGAGAGGCCACTGGA 120
QY	797 TCGGCC TCACAG ACTCAG AGCGTGA AATGAAGTGGCTGGATGGGACATCTCCAG 856

COMMENT	Contact: Robert Strausberg, Ph.D. Email: rcapbs-r@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM9128 row: m column: 13 High quality sequence stop: 580.									
	Location/Qualifiers 1. .580 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:3962292" /clone_lib="NCI_CGAP_Maml" /tissue_type="tumor, biopsy sample" /dev_stage="10 months, virgin" /lab_host="DH10B" /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"									
	BASE COUNT 159 a 153 c 161 g 107 t									
	ORIGIN									
	Query Match 32.9%; Score 445.6; DB 12; Length 580; Best Local Similarity 85.5%; Pred. No. 1.3e-80; Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;									
	Qy	290	GCGGCAAGGATCTAAAGGCTCCAGGGCCCCAAAGGCTCCGCTGTTCCCTGGGAAGC	349						
	Db	1	GCAGCAAGGATCCAAAGGCTCACAGGTCCCAAGGATCTGTTGGTCCCGAGGAAGC	60						
	Qy	350	CGGGCCCTCAGGGCCCCAGTGGGGACCAGGCCCCCGGGGCCACAGGCAAGAGGAC	409						
	Db	61	CTGGCCCTCAAGGACTAGTGGGGACCCAGGACCACAGGTCCACAGGAGGATGGAC	120						
	Qy	410	TCCCGCCCTCAGGGCCCTCTGCTTCCAGGGACTTCAGGGACCCGTGGGAGCCTG	469						
	Db	121	TCCCTGGCCCTCAGGGCCCTCTGCTTCCAGGGACTACAGGGCACTGTGGTGAGCCTG	180						
	Qy	470	GGTGCTCTGAGCCCTCGGGGACTGCCAGGCTTGCTGGGGTACAGGCATGCCAGGCCCA	529						
	Db	181	GAGTACTTGAGCTCGGGGTGCCAGGCTGCCAGGGTGCCAGGCATGCCGTGGCCCTA	240						
	Qy	530	AGGGCCCCCGGCCCTCTGGCCCATCAGGAGCGGTGGTGCCTCGCCCTGCAGAAATG	589						
	Db	241	AGGGACCACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCAATGGCTCTGCAGAAATG	300						
	Qy	590	AGCCAAACCCGGCAGCAGGACATGCTGCCGCCCTCACTGGGAAGAACTTCACAGACA	649						
	Db	301	AACCAACCCAGCATCAGAGGTCAACGGATGTCGCCCTCACTGGGAAGAACTTCAGATA	360						
	Qy	650	AATGCTACTATTTTTCAGTTGAGAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAG	709						
	Db	361	AATGCTACTATTTTTCATTGGAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAG	420						
	Qy	710	ACAAGCTTCACATCTTTTTCATAAACAACACTAGAGAGGAACAGCAATGGATAAAAAAC	769						
	Db	421	ACAAATCTTCCCATCTCGTTTTCATAAACAACACTCAAGAGAAGACAGCAATGGATAAAAAAGC	480						
	Qy	770	AGATGGTAGGAGAGAGCCACTGGATCGGCCTCACAGACTCAGAGCGTGAATATGAAT	829						
	Db	481	ATACCGTGGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGAAACAGGAAACCGAAT	540						
	Qy	830	GGAAGTGGCTGGATGGACATCTCCAGACTACAAAAATTG	869						
	Db	541	GGAAGTGGCTAGACGGGTCACTGTTGATACAAAACTG	580						

RESULT 12 BQ771366 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT	BQ771366 808 bp mRNA linear EST 26-JUL-2002 UI-M-F10-byu-g-09-0-UI-r1 NIH_BMAP_F10 Mus musculus cDNA clone IMAGE:5702432 5', mRNA sequence. BQ771366 BQ771366 EST. BQ771366.1 GI:21979842 house mouse. Mus musculus. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rcapbs-r@mail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP)									
	Seq primer: pYX-5. Location/Qualifiers 1. .808 /organism="Mus musculus" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:5702432" /clone_lib="NIH_BMAP_F10" /tissue_type="whole brain" /dev_stage="embryo 12.5dpc" /lab_host="DH10B (T1 phage resistant)" /note="Organ: Brain; Vector: pYX- Asc; Site:1: EcoR I; Site:2: Not I. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCACGAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."									
	BASE COUNT 270 a 158 c 173 g 206 t 1 others									
	ORIGIN									
	Query Match 32.6%; Score 441.8; DB 14; Length 808; Best Local Similarity 77.5%; Pred. No. 7.6e-80; Matches 604; Conservative 0; Mismatches 157; Indels 18; Gaps 5;									
	Qy	576	GGCCCTGCAGATGAGCCAAACCCCGGACCGGAGGACAAATGGCTGCCGCCCTCACTGGAA	635						
	Db	1	GGCTCTGCAGATGAACCAACCCAGCATCAGAGGTCAACGGATGTCGCCCTCACTGGAA	60						
	Qy	636	GAACCTTCAGACAAATGCTACTATTTTTCAGTTGAGAAGAAATTTTGGAGATGCTCAA	695						
	Db	61	GAACCTTCAGATAAATGCTACTATTTTTCATTGGAAAAAGAAATTTTGGAGATGCTCAA	120						
	Qy	696	GCTTTTCTGTGAAGACAACTCTTCACATCTCTTTTTCATAAACAACACTAGAGAGAAACAGCA	755						
Db	121	GCTTTTCTGTGAAGACAAATCTTCCCATCTCGTTTTCATAAACAACACTCAAGAGAGACAGCA	180							
Qy	756	ATGGATAAAAAACAGATGGTAGGGAGAGAGAGCCACTGGATCGGCCTCACAGACTCAGA	815							

```
|||||
Db 181 ATGATAAAAAACATACCGTGGGAGAGAAAGCCATTGGATCGGCTCACAGACTCAGA 240
QY 816 GCGTGAATGAATGGAAGTGGTGGGAGCATCTCCAGACTACAAAAATTGGAAGC 875
Db 241 ACAGGAACGAATGGAATGGCTAGAGGGTGCACCTGTGATTACAAAACTGGAAGC 300
QY 876 TGACAGCGCGGTAATCTGGGGTCATCGGCCATGGCCAGGAGAGACTGTGCTGGGTGAT 935
Db 301 TGGACACCAAGATACTGGGGCAGTGGCCATGGCCAGGAGAGACTGTGCTGGCTGAT 360
QY 936 TTATGCTGGCAGTGAACGATTTCAATGTGAAGACGTCAATAACTTCATTTGGGAAA 995
Db 361 TTACGAGGACAGTGAATGACTTCCAGTGTGATGAATCAATAACTTCATTTGTGAGAA 420
QY 996 AGACAGGGAGACACTCTCATCTGCATTATTAACGGACTGTGATGGGATCATGAGCA 1055
Db 421 GGAAGGGAGGAGTACCATCATCATTAATAACAGCATGATATAANTAGAGAAACATA 480
QY 1056 AATTTTCA--GCTCTCAAAAGGCAAGACACTCCTTTCTTAATTGCATCACCTTCTCATCA 1113
Db 481 TTTTCTGATGCTCTGAAAGCCGAAGATGCTCGTTTTTTGATTCCATCA-CITCTCACCA 539
QY 1114 GATTGAAAAAAGAACGACTGAAAAACCAATTAATCTGAAAAAATTAAGACAGTGTGT 1173
Db 540 GATTG--AATGAAAAAGCTCTGAAAAGTAGTATTCAAAATAAATGGAACACTACTG-- 595
QY 1174 TTTTACCATCGTCAATCCCAAGACTTGGGAACCTAAATG-TTCCCCAGGCGTATATG 1232
Db 596 -----CACATAACCCAGGACTAGGGGCTTAAATGCTCCCCAAGTTGATATA 645
QY 1233 CTGATTTTCATTTGTCACATGGAATGACATAGATTCTCTCGTCAGTAACCGTGC 1292
Db 646 TTGATTTCCAGTGTCAAAATGAGTGAATCGCATAGATTTTCTCAGCCATTAACCATAGA 705
QY 1293 GATTATACAAATATGCTTCCAAAGTATGGAACACTCAATCAGAAAAAGGTTATCAT 1351
Db 706 ATTTATGCAAGATATATCTTTCCAAATATGGAATGCTCCCAATCAGAAAAAGACTATCAT 764

RESULT 13
AL543000/c
LOCUS
DEFINITION AL543000 LTI_FL002_P11 Homo sapiens cDNA clone CS0DE013Y08 3 prime
, mRNA sequence.
ACCESSION AL543000
VERSION AL543000.1 GI:12875478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 517)
Li,W.B., Gruber,C., Jessee,J. and Pollayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE013Y08"
/lab_host="LTI_FL002_P11"
/Note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
```

```
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 70 a 117 c 117 g 201 t 12 others
ORIGIN
Query Match 30.8%; Score 416.8; DB 9; Length 517;
Best Local Similarity 89.5%; Pred. No. 9,3e-75;
Matches 451; Conservative 7; Mismatches 45; Indels 1; Gaps 1;
QY 729 TTTTCATAAACAACACTAGAGAGGACCAATGGATAAAAAACAGATGGTAGGAGAGAGAG 788
Db 517 TTTTCATAAACAACACTAGAGAGGACCAATGGATAAAAAACAGATGGTAGGAGAGAGAG 458
QY 789 CCACATGGATCGGCCCTCACAGACTCAGAGCGTCAAAATGAATGAAATGGCTGGTGGGAC 848
Db 457 CCACATGGATCGGCCCTCACARACTCAGAGCGTCAAAATGAATGAAATGGCTGGTGGGAC 398
QY 849 ATCTCCAGACTACAAAAATTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGG 908
Db 397 ATCTCCAGACTACAAAAATTGAAAGCTGGACAGCCGGATWACTGGGGTCATGGCCATGG 338
QY 909 GCACAGAGAAGACTGTGCTGGGTTCATTTATGCTGGCAGTGAAGGATTTTCCATGTGA 968
Db 337 GCACAGAGAAGACTGTGCTGGGTTCATTTATGCTGGCAGTGAAGGATTTTCCATGTGA 278
QY 969 AGACGCTCAATAACTTTCATTTGCGAAAAACAGAGAGAGACTGTCTATCTGCATTATA 1028
Db 277 AGACGCTCAATAACTTTCATTTGCGAAAAACAGAGAGAGAGACTGTCTATCTGCATTATA 218
QY 1029 ACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAGGCAAAAGGACACTCT 1088
Db 217 ACGGACTGTGATGGGATCACATGAGCAAAATTAACAGCMCAAAAGGCAAAAGGACACCCC 158
QY 1089 TTCTAATTCATCACCCTTCTCATCAGATTGAAAAAAGGAGAGAGAGAGAGAGAGAGAG 1148
Db 157 TCCTAATTCATCACCCTTCTCATCAGATTGAAAAAAGGAGAGAGAGAGAGAGAGAGAGAG 99
QY 1149 TGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCGCTCATTTACCAAGAGCTTGGGAAC 1208
Db 98 CGAAAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 39
QY 1209 TAAATGTGTTCCCGAGGGTGATATG 1232
Db 38 CAAAGCGCCCGCCAGGGCGGAAACG 15

RESULT 14
BE290299
LOCUS
DEFINITION BE290299 601 bp mRNA linear EST 26-OCT-2000
601089246F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5',
mRNA sequence.
ACCESSION BE290299
VERSION BE290299.1 GI:9171250
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 601)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

Plate: L1AM8516 row: k column: 12
High quality sequence stop: 551.
Location/Qualifiers

FEATURES
source

1. .601
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3484163"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 156 a 161 c 179 g 105 t

Query Match 30.7%; Score 415.8; DB 10; Length 601;
Best Local Similarity 84.8%; Pred. No. 1.5e-74;
Matches 501; Conservative 0; Mismatches 87; Indels 3; Gaps 3;

Qy 196 AAGGACACAAAGGAGAGAGAGGGAGCGCTGGACACACTGCGCCCTCGGGTGAGAGAGCC 255
Db 9 AGGTGACAGAAAGGAGAGAGAGAGAGCGCTGTCACCTGCGCCCTCGGGTGAGAGAGGCC 68
Qy 256 CCAATTGACACAGCTGGTCCCGCCGAGAGAGCGTGGCGCAAGGATCTAAAGGCTCCACAG 315
Db 69 ACAATTGACACAGCTGGCGCCCTCGGTGAGAGCGTGGCGCAAGGATCTAAAGGCTCCACAG 128
Qy 316 GGCCCCAAAGGCTCCCGTGTTCCTCGG-AAGCGCCGCGCCCTCAGGGCCCCAGTGGGGA 374
Db 129 GGTCCCAAGGATCTCGTGGTCCCGAGGTAAAGCTGACCTCAAGGACCTAGTGGGGA 188
Qy 375 CCGAGGCCCGCGGCCCGCCAGGACGAAAGAGGAGTCCCGCGCCCTCAGGGCCCTCTCTGG 434
Db 189 CCCAGGACACAGGTCCAGGCAAGGATGGACTCCCTGCGCCCTCAGGGCCCTCTCTGG 248
Qy 435 CTTCCAGGACTTTCAGGCGACCGTTGGGAGCGCTGGGGTGGCTTGACCTCGGGGACTGCC 494
Db 249 CTTCCAGGACTTACAGGCGACTGTGGTGAGCGTGGAGTACTTGACCTCGGGGTTGCC 308
Qy 495 AGGCTGCTGGGTGACAGGCGATGCCAGGCCCAAGGCCCGCCCGCCCTCTCTGGGCC 554
Db 309 AGGCTTGCC-AGGCTGCGCAGCATGCTGGGCGCTTAAGGACCACTGCGCCCTCAGGCC 367
Qy 555 ATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCACCCCGGACCGGAGGACAA 614
Db 368 CTCAGGAGCAATGGAGCCATTGGCTCTGCAGATGAACCAACCCCGCAGCATCAGAGGTCAA 427
Qy 615 TGGCTGCGCGCTCACTGGAAGAACTTCACAGCAAAATGCTACTATTTTTCAGTTGAGAA 674
Db 428 CGATGTCGCGCTCACTGGAAGAACTTCACAGATAAATGCTACTATTTTTCATTGGAAGA 487
Qy 675 AGAATTTTGGAGTCAAAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTTCAT 734
Db 488 AGAA-TTTTGAAGATGCTAAGCTTTCTGTGAAGACAAATCTTCCCATCTCGTTTTCAT 546
Qy 735 AAACACTAGAGAGGAACAGCAATGATGATATAAAACAGATGATGAGGAGAGA 785
Db 547 AAACACTAGAGGAACAGCAATGGGTAAAAAAGCATACGTGGGAGAGA 597

RESULT 15
BB248064
LOCUS
DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus
musculus cDNA clone A730023E20 3', mRNA sequence.
ACCESSION BB248064
VERSION BB248064.2 GI:16355610
KEYWORDS EST.

SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 638)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 6, 2000 this sequence version replaced gi:8940810.

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URU:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,K.I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES
source

Location/Qualifiers

1. .638
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730023E20"
/clone_lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCAAGAGCTCTTTTTTTTTTTTTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTAAATTAATCCGCCGCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pbluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT	212 a	134 c	146 g	146 t
ORIGIN				
Query Match	30.0%	Score 406;	DB 10;	Length 638;
Best Local Similarity	80.4%	Pred. No. 1.5e-72;		
Matches 513; Conservative	0;	Mismatches 120;	Indels	5; Gaps 3;
QY	536	CCCCCGGCCCTCTGGCCCTCAGGAGGGTGGTGGCCCTGCAGAAATGAGCCAA	595	
Db	1	CACTTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCATTGGCTCTGCAGAAATGACCAA	60	
QY	596	CCCCGGCACCGGAGGACAAATGGCTGCCGCCCTCACTTGGAAAGAACTTCACAGACAAATGCT	655	
Db	61	CCCCAGCATCAGAGGTCAACGGATGTCGCCCTCACTGGAAAGAACTTCACAGATAAATGCT	120	
QY	656	ACTATTTTTCAGTTCAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGT	715	
Db	121	ACTATTTTTCATGGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAAT	180	
QY	716	CTTCACATCTGTTTTCATAAACACTAGAGAGGACAGCAATGGATAAAAAACAGATGG	775	
Db	181	CTTCCCATCTCGTTTTCATAAATCAAGAGAGACAGCAATGGATAAAAAGCATACCG	240	
QY	776	TAGGAGAGAGAGCCACATGGATCGGCCCTCACAGACTCAGAGCGTGAAAATGAATGGAAGT	835	
Db	241	TGGGAGAGAGAAAGCCATTGGATCGGCCCTCACAGACTCAGAACAGGAAGCGAATGGAAGT	300	
QY	836	GGCTGGATGGGACATCTCCAGACTACAAAAATTGAAAGCTGCACAGCCGGATAACTGGG	895	
Db	301	GGCTAGACGGGTACCTGTGATTACAAAACCTGAAAAGCTGGACACACAGATAACTGGG	360	
QY	896	GTCATGGCCATGGCCAGGAGAACTGTGCTGGGTGATTTATGCTGGGCAGTGGAAAG	955	
Db	361	GCAGTGGCCATGGCCAGGAGAACTGTGCTGGCTGATTTACGACAGACAGTGAATG	420	
QY	956	ATTTTCAATGTGAAGACGTCAATACTTCATTTGGGAAAAGACAGGGAGACAGTACTGT	1015	
Db	421	ACTTCCAGTGTGATGAATACTCAATACTTCATTTGTGAGAAAGGAGGGAGGCAGTACCAT	480	
QY	1016	CATCTGCATTATAACGGACTGTGATGGGATCACATGACAAATTTTCA--GCTCTCAAG	1073	
Db	481	CATCCATATTATACAGCATGATATAATAGCAGAAACATATTTCTGATGCGCTCTGAAG	540	
QY	1074	GCAAGGACACTCTTTCTTAATTGATCACCCTTCTCATCAGATTGAAAAAAGCA	1133	
Db	541	CGAAGAATGCTCGTTTTTGTATTCATCA-CITCTCACCAGATTG--AATGGAAAAAGCT	597	
QY	1134	CTGAAACCAATTACTGAAAAAATTTGACAGCTAGT	1171	
Db	598	CTGAAAGTAGTTATTCAAAATAAATGGACACCTACTG	635	

Search completed: March 21, 2003, 08:08:35
Job time : 1791.64 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 02:20:06 ; Search time 286.911 Seconds
(without alignments)
4293.470 Million cell updates/sec

Title: US-09-763-712A-2

Perfect score: 2969

Sequence: 1 MYGHNVMNLLNLTVQV.....EDVNNFICKRETVLSSAL 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=Cgn2_1/USPTO_spool/US09763712/runat_14032003_100949_18091/app_query.fasta_1.1877
-DB=N_Geneseq_101002 -QFMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.qcgn_1.1.626 -runat_14032003_100949_18091 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	2969	100.0	2024	21	AAA07697	Human collectin en
2	2969	100.0	2262	24	ABA97932	Human scavenger re
3	2969	100.0	2628	22	AAH43036	Nucleotide sequenc
4	2963	99.8	2181	22	AAI60628	Human polynucleoti
5	2963	99.8	2641	22	AAC66903	Human EXMAD-14 cod
6	2956	99.6	2318	22	AAI58842	Human polynucleoti
7	2936	98.9	2929	24	ABQ92072	Human polynucleoti
8	2936	98.9	2930	20	AAV55746	Human secreted pro
9	2765	93.1	2637	22	AAH43037	Nucleotide sequenc
10	2688	90.5	2005	24	ABL95574	Human angiogenesis
11	2688	90.5	2005	24	ABL88085	Human PRO7223 cDNA
12	2187	73.7	2256	22	AAH43054	Nucleotide sequenc
13	1715.5	57.8	1521	23	AA571133	DNA encoding novel
14	765	25.8	3685	20	AAZ7858	Human CSR3 protein
15	765	25.8	3810	20	AAZ7856	Human CSR1 protein
16	474.5	16.0	1410	24	ABK84622	Human cDNA differe
17	474.5	16.0	1410	24	ABL66964	Lung cancer relate
18	471	15.9	5676	15	AAQ64556	Human collagen (Ty
19	471	15.9	8284	22	AAAL26526	Human breast cance
20	471	15.9	8284	22	AAAL26553	Human breast cance
21	471	15.9	8284	22	AAAL26600	Human breast cance
22	462.5	15.6	4428	22	AAD06574	Bovine alphas(III)
23	462.5	15.6	4428	22	AAD06575	Bovine alphas(III)
24	459.5	15.5	756	14	AAQ43034	Collagen-like poly
25	459.5	15.5	756	17	AAT16768	Collagen-like poly
26	458.5	15.4	4821	23	AA586866	DNA encoding novel
27	454	15.3	3171	21	AAAL12503	DNA encoding a hu
28	454	15.3	3174	21	AAZ99843	DNA encoding human
29	453	15.3	6158	24	ABL62095	Colon adenocarcino
30	453	15.3	6158	24	ABL65452	Lung cancer relate
31	453	15.3	6158	24	ABK35486	Human endometrial
32	452.5	15.2	3170	21	AAAL12493	cDNA encoding huma
33	452.5	15.2	3171	21	AAAL12502	cDNA encoding a hu
34	452.5	15.2	3181	19	AAV59358	Nucleotide sequenc
35	452.5	15.2	3349	17	AAT16518	Collagen A1/decori
36	452.5	15.2	3349	21	AAAL12500	cDNA encoding a ch
37	452.5	15.2	3531	21	AAAL12497	cDNA encoding a ch
38	452.5	15.2	3535	17	AAT16515	Collagen A1/BMP-2B
39	452.5	15.2	3541	17	AAT16516	Collagen A1/TGF-be
40	452.5	15.2	3541	21	AAAL12498	cDNA encoding a ch
41	452.5	15.2	4192	17	AAT16517	Collagen A1/decori
42	452.5	15.2	4409	19	AAV60814	Human recombinant
43	452.5	15.2	4428	22	AAD06578	Porcine alphas(III)
44	452.5	15.2	4770	22	AA522441	Human cDNA encodin
45	452.5	15.2	6728	22	AAF90491	Human pro-alpha-1

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX AAA07697;
AC AAA07697;
XX 20-JUN-2000 (first entry)
DT 20-JUN-2000 (first entry)
XX Human collectin encoding DNA.
DE Human collectin encoding DNA.
XX Collectin; human; antibacterial; antiviral; ds.
KW Collectin; human; antibacterial; antiviral; ds.
OS Homo sapiens.
XX Location/Qualifiers
FH Key
FT CDS
FT 55..1698
FT /*tag= a
FT /product= "collectin"

FT	misc_feature	1..738	
FT		/*tag= b	
FT		/note= "fragment specifically claimed in claim 8"	
FT	misc_feature	55..738	
FT		/*tag= c	
FT		/note= "fragment specifically claimed in claim 8"	
FT	misc_feature	79..738	
FT		/*tag= d	
FT		/note= "fragment specifically claimed in claim 8"	
FT	misc_feature	325..738	
FT		/*tag= e	
FT		/note= "fragment specifically claimed in claim 8"	
FT	misc_feature	358..738	
FT		/*tag= f	
FT		/note= "fragment specifically claimed in claim 8"	
FT	misc_feature	670..1695	
FT		/*tag= g	
FT		/note= "fragment specifically claimed in claim 5"	
FT	misc_feature	685..738	
FT		/*tag= h	
FT		/note= "fragment specifically claimed in claim 7"	
FT	misc_feature	730..738	
FT		/*tag= i	
FT		/note= "fragment specifically claimed in claim 7"	
FT	misc_feature	739..1695	
FT		/*tag= j	
FT		/note= "fragment specifically claimed in claim 6"	
FT	misc_feature	1696..2024	
FT		/*tag= k	
FT		/note= "fragment specifically claimed in claim 9"	
XX			
XX	WO200011161-A1.		
XX			
XX	02-MAR-2000.		
XX			
PF	24-AUG-1999;	99WO-JP04552.	
XX			
PR	24-AUG-1998;	98JP-0237611.	
XX			
XX	(FUSO) FUSO PHARM IND LTD.		
PA			
XX			
PI	Wakamiya N;		
XX			
DR	WPI; 2000-224696/19.		
DR	P-PSDB; AAY77985.		
XX			
XX	New collectin of human origin having antibacterial and antiviral		
PT	activity, and gene encoding it useful for production of transgenic		
PT	animals and of antibodies for screening potential drug molecules -		
XX			
PS	Claim 5-9; Page 87-90; 106pp; Japanese.		
XX			
CC	The invention relates to polynucleotides encoding a new collectin of		
CC	human origin. The collectin can be used as an antibacterial and antiviral		
CC	agent and for screening potential drug molecules. The new collectin can		
CC	be produced by standard recombinant methodology. The present sequence		
CC	represents a DNA encoding the human collectin.		
XX			
SQ	Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;		
	Alignment Scores:		
	Pred. No.:	5.17e-96	Length: 2024
	Score:	2969.00	Matches: 547
	Percent Similarity:	100.00%	Conservative: 0
	Best Local Similarity:	100.00%	Mismatches: 0
	Query Match:	100.00%	Indels: 0
	DB:	21	Gaps: 0
US-09-763-712A-2 (1-547) x AAA07697 (1-2024)			
Qy	1 MetTySerHisAsnValVallleMetAsnLeuAsnLeuAsnLeuThrGlnValGln 20		
Db	55 ATGTATTCTCAATGTGTCATGACCTCAACACCTGAACCTGACCGGTGCAG 114		

QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 Db 1195 CCCAAGGCCCCCGCCCTCTGGCCCATCAGAGCGGTGGTCCCTGGCCCTGGAG 1254
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 Db 1255 AATGAGCCACCCCGCCACCGAGGACAAATGGCTGCCCGCTCACTGGAGAAGCTTCACA 1314
 QY 421 AspLysCysTrpTyrPheSerValGluLysGluLePheGluAspAlaLysLeuPheCys 440
 Db 1315 GACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTGT 1374
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
 Db 1375 GAAGACAAAGTCTTCACATCTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAA 1434
 QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 Db 1435 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCCTCAGACTCAGAGCGTGAAAT 1494
 QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
 Db 1495 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCG 1554
 QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
 Db 1555 GATAACTGGGTCACTGGCCATGGCCAGGAGAAAGACTGTGCTGGTTGATTATCTGGG 1614
 QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
 Db 1615 CAGTGGACAGATTCCAAATGGAAGACGTCATTAATCTTTCGAAAGACAGGGAG 1674
 QY 541 ThrValLeuSerSerAlaLeu 547
 Db 1675 ACAGTACTGTCTATCGCATTA 1695
 RESULT 2
 ID ABA97932 standard; cDNA; 2262 BP.
 AC ABA97932;
 XX 25-APR-2002 (first entry)
 DT Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
 DE Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
 KW rheumatoid arthritis; gene; ss.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH 33..2262
 CDS /*tag= a
 FT /product= "scavenger receptor-like protein"
 XX JP2001340089-A.
 PN 11-DEC-2001.
 PD 08-DEC-2000; 2000JP-0375066.
 PF 27-MAR-2000; 2000JP-0090772.
 PR (SHIO) SHIONOGI & CO LTD.
 PA WPI; 2002-144965/19.
 DR P-PSDB; ABB08642.
 XX New scavenger receptor-like protein for diagnosis, prevention and
 PT treatment of autoimmune disease, such as rheumatoid arthritis -
 XX Claim 3; Fig 1; 38pp; Japanese.

XX The invention relates to a human scavenger receptor-like protein. The
 CC protein is useful as a target molecule for diagnosis, prevention and
 CC treatment of autoimmune diseases such as rheumatoid arthritis.
 SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
 Alignment Scores:
 Pred. No.: 5.63e-96 Length: 2262
 Score: 2969.00 Matches: 547
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps:
 US-09-763-712A-2 (1-547) x ABA97932 (1-2262)
 QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 Db 619 ATGTAATCTCATAATGTGGTCAATCATGAACCTCAACAACCTGAACCTGACCAGGTGCAG 678
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
 Db 679 CAGAGAACTCATCCAGAAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCCAG 738
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 739 CGAATCAAGAAAGCGACTTTCAAAATCTGCAGCAGGTTTTTCTCAAGCCAAAGAGACAG 798
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 Db 799 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGACCTGGCTGCCAACACTCTCGCTTG 858
 QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 859 GCCAAAGCCAAACAGACACACCTGGAGGATATGACAGCCAGCTCACTCATTCACAGT 918
 QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 919 CAGATGGAGAACATCACCACATATCTCAAGCCAAAGCAGACAGCAACTGAAGACCTGCAG 978
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 979 GACTTACACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCACTGGAGGAACGC 1038
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
 Db 1039 TTCCAGCTTTTGAGACCGGATATTGTGAACATCATTAGCAATATCAGTTACACGCCAC 1098
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 1099 CACCTGCGGACGCTGACCAGCAATCTAAATGAAGTCAGGACCCTGCACAGATACCCCTT 1158
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 1159 ACCAAACACACAGATGATCTGACCTCTCTGAATAATACCTGGCCAAACATCGCTTGGAT 1218
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 1219 TCTGTTTCTCAGGATGCACAAAGATTGTGAGGTGAGGTTGACACTGAAGTAGGCC 1278
 QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 Db 1279 AACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAGCATGGTCAGCTCATC 1338
 QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
 Db 1339 AAGAAATTTTACAATACTACAAGGTCCACGGGCCCGGGTCCAAGAGGTGACAGAGGA 1398
 QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 Db 1399 TCCAGGGACCCCTGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGAGGCT 1458

QY	281	GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu	300
Db	1459	GGACCACTGGCCCTCGGGCTGAGAGAGGCCCAATTTGGACCACTGGTCCCGCCGAGAG	1518
QY	301	ArgGlyGlySerGlySerGlySerGlnGlyProLysGlySerArgGlySerProGly	320
Db	1519	CGTGGCGGCAAGAGATCTAAAGGCTCCCAAGGGCCCAAGGCTCCGTTCCCTGGG	1578
QY	321	LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProProGlyProGlyLysGlu	340
Db	1579	AAGCCCGCCCTCAGGCCCCAGTGGGACCCAGGCCCGCCCGGGCCCAAGCAAGAG	1638
QY	341	GlyLeuProGlyProGlnGlyProProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu	360
Db	1639	GGACTCCCGGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAG	1698
QY	361	ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly	380
Db	1699	CTGGGGTGGCTGGACCTCGGGACTGCCAGGCTTGCCCTGGGGTACCAGGCATGCCAGGC	1758
QY	381	ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln	400
Db	1759	CCCAAGGGCCCCCGCCCTCTGGCCCATCAGGAGCGTGGTCCCTGGCCCTGCAG	1818
QY	401	AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr	420
Db	1819	AATGAGCCAAACCGGCACGGAGGACAATGGTGGCCCTCACTTGAAGAAGACTTCACA	1878
QY	421	AspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys	440
Db	1879	GACAAATGCTACTATTTTCAGTTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGT	1938
QY	441	GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys	460
Db	1939	GAAGACAAGTCTTCACATCTGTTTTCATAAACACTAGAGAGAACAGCAATGGATAANA	1998
QY	461	LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn	480
Db	1999	AAACAGATGGTAGGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAANA	2058
QY	481	GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyLysPro	500
Db	2059	GAATGGAATGGCTGGATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCG	2118
QY	501	AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly	520
Db	2119	GATAACTGGGGTCATGGCCATGGCCAGGAGAAAGACTGTGCTGGGTGATTTATGCTGG	2178
QY	521	GlnTrpAsnAspPheGlnCysGluAspValAlaAsnAsnPheIleCysGluLysAspArgGlu	540
Db	2179	CAGTGGAGACGATTCCAATGTGAAGACGTCAATAACTTCATTGCGAAAAAGACAGGGAG	2238
QY	541	ThrValLeuSerSerAlaLeu	547
Db	2239	ACAGTACTGTCATCTGCATTA	2259
RESULT 3			
ID	AAH43036		
XX	AAH43036 standard; DNA; 2628 BP.		
AC	AAH43036;		
DT	15-OCT-2001 (first entry)		
DE	Nucleotide sequence of a human scavenger receptor.		
KW	Human; scavenger receptor; SRCL-PI; macrophage; basal immunity;		
KW	arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;		
KW	low density lipoprotein; ss.		
OS	Homo sapiens.		
Key	Location/Qualifiers		

FT	CDS	74..2302	
FT		/tag= a	
XX		/product= "scavenger receptor"	
PN	WO200159107-A1.		
XX	16-AUG-2001.		
XX	08-FEB-2001; 2001WO-JP00874.		
PR	14-FEB-2000; 2000JP-0035155.		
PR	10-OCT-2000; 2000JP-0309068.		
PA	(FUSO) FUSO PHARM IND LTD.		
PI	Wakamiya N;		
DR	WPI; 2001-497076/54.		
DR	P-PSDB; AAG63346.		
XX	New Scavenger receptor proteins SRCL-PI with collectin-like structure, useful for treatment and diagnosis of diseases associated with oxidized low-density lipoprotein accumulation		
PS	Claim 2; Page 79-84; 118pp; Japanese.		
CC	The present sequence encodes a human scavenger receptor, designated SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They are useful in clarifying the functions of macrophages and basal immunity. They are also useful in the treatment, prevention, diagnosis and investigation of diseases such as arteriosclerosis, diabetic complications, bacterial infection and restenosis following angioplasty, CC which are associated with accumulation of oxidized low density lipoprotein and the binding of advanced glycation end-products into cells.		
SQ	Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;		
Alignment Scores:			
Pred. No.:	6,31e-96	Length:	2628
Score:	2969.00	Matches:	547
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0
US-09-763-712A-2 (1-547) x AAH43036 (1-2628)			
QY	1	MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuThrGlnValGln	20
Db	659	ATGTATTCTCATAATGTGGTTCATCATGAACCTCAACAACCTGAACCTGACCGGTGCAG	718
QY	21	GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln	40
Db	719	CAGAGAACCTCATCAGGAATCTGACGGTCTGTGGATGACACAGCCAGGCTATCCAG	778
QY	41	ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr	60
Db	779	CGAATCAAGACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGGACACG	838
QY	61	AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu	80
Db	839	GATTGGCTGAAGGAGAAAGTCAGAGCTTGCAGACGCTGGCTGCCAACAACTCTCGGTTG	898
QY	81	AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly	100
Db	899	GCCAAGGCCAACAGCACACCTGGAGGATATGACAGCCAGCTCAACTCATTCACAGT	958
QY	101	GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln	120
Db	959	CAGATGGAGAACATCACCACCTATCTCTCAAGCCACGACAGCACTGAAAGACCTGCAG	1018
QY	121	AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg	140

Db 1019 GACTTACCAAGATGACAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGAACGC 1078
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleLeuSerAsnIleSerTyrThrAlaHis 160
Db 1079 TTCACAGCTCTTTGACAGCGATATGTTGACATCATATTACATATCATGTTACACAGCCAC 1138
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1139 CACCTGCGGAGCGTGACAGCAATCAAAATGAAGTCAGGACCACTTGCACAGATACCCCTT 1198
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1199 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCCGTTTGAT 1258
Qy 201 SerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1259 TCTGTTCTCTCAGAGTGCACACAGATTTGATGAGTCGAGGTAGACACTGAAGTAGCC 1318
Qy 221 AsnLeuSerValIleMetGluGlnMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1319 AACTTATCAGTCATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1378
Qy 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
Db 1379 AAGAAATTTTACAATACTACAAGGTCCACCGGCCCCAGGGGTCCAAGAGGTGACAGAGA 1438
Qy 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
Db 1439 TCCAGGAGACCCCTGGCCCACTGGCAACAGGAGACAGAAGAGAGGGGGAGGCT 1498
Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
Db 1499 GGACCACTGGCCCTGCGGCTGAGAGAGGCCAATTTGGACAGCTGGTCCCCCGGAGAG 1558
Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1559 CGTGGCGCAAGAGGATCTAAAGGCTCCAGGCGCCCAAGGCTCCCGTGTCCCTCTGGG 1618
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
Db 1619 AAGCCCGGCTTCAGGCGCCCATGTTGGGAGCCAGGCCCCCGGCCCCACAGGAAAAG 1678
Qy 341 GlyLeuProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1679 GGACTCCCGGCGCTCAGGCGCTCTGCTGCTCCAGGACITTCAGGGCACCGCTTGGGAG 1738
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1739 CCTGGGTCCTGGACCTCGGGGACTGCGAGGCTTGCCTGGGGTACCAGGCTGCCAGGC 1798
Qy 381 ProLysGlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
Db 1799 CCCAAGGCGCCCGCGGCTCTCTGGCCCATCAGAGCGGTGGTCCCTCCCTCGCCCTGGCAG 1858
Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db 1859 AATGAGCAACCCCGGACCGGAGACAATGGCTCCCGCTCACTGGAAGAACTTCACA 1918
Qy 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1919 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGT 1978
Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db 1979 GAAGACAGTCTTCACATCTTGTGTTTTCATAAACACTAGAGGAGACAGCAATGGATAAAA 2038
Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2039 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCCTCACAGACTCACAGCGTGAAT 2098
Qy 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500

Db 2099 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTTCAAAAAATTGGAAGACTGGACAGCG 2158
Qy 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2159 GATACTGGGTTCATGGCCATGGGCGAGAGAACTGTGCTGGTGTGATTTATGCTGGG 2218
Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2219 CAGTGAACGATTTTCCAATGTGAAGACGTCATAACTTTCATTGGCGAAAAGACAGGAG 2278
Qy 541 ThrValLeuSerSerAlaLeu 547
Db 2279 ACAGTACTGTCTATCTGCATTA 2299
RESULT 4
ID AAI60628 standard; cDNA; 2181 BP.
XX AAI60628;
AC AAI60628;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4617.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.
KW
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41472.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4617; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;

Alignment Scores:

Pred. No.: 8.86e-96 Length: 2181
 Score: 2963.00 Matches: 546
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 99.80% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712a-2 (1-547) x AAI60628 (1-2181)

QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 DB 47 ATGTATTCTCATATGTTGGTTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAG 106
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 DB 107 CAGAGGAACCTCATCAGGAATCTGCAGCGTCTGTGGATGACACAAGCCAGGCTATCCAG 166
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 DB 167 CGAATCAAGAACACACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGGACAG 226
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 DB 227 GATTGGCTGAAGGAGAAAGTGACAGCTTCGACACACTGGCTGCCAACAACCTCGCGTTG 286
 QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 DB 287 GCCAAAGCCCAACACACACCTCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 346
 QY 101 GlnMetGluAsnIleThrPheSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 DB 347 CAGATGGAGAACATCACCACTATCTCTCAAGCCACAGCAGCAACCTGAAAGACCTGCAG 406
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 DB 407 GACTTACACAAAGATGCAGAGATAGACAGCCATCAAGTTCAACCACTGGAGGAACCC 466
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleLeSerAsnIleSerThrAlaHis 160
 DB 467 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACAGCCAC 526
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 DB 527 CACCTGGCGACCGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACAGATACCCCT 586
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 DB 587 ACCAAACACACAGATGATCTGACCTCTTGAATATACCTGGCCCAACATCCGTTTGGAT 646
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 DB 647 TCTGTTTCTCTCAGGATGCACAAAGATTGATGAGTGCAGGTTAGACACTGAAGTAGCC 706
 QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 DB 707 AACTTATCAGTGATTATGGAGAAATGAAGTAGTAGACTCCAGCATGGTTCAGCTCATC 766
 QY 241 LysAsnPheThrIleLeuGlnGlnProGlyProArgGlyProArgGlyAspArgGly 260
 DB 767 AAGAATTTTACAATACTACAAGGTCCACCGGCCCGCCAGGGTCCCAAGAGGTGACAGAGGA 826
 QY 261 SerGlnGlyProGlyProThrClyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 DB 827 TCCAGGACCCCTGGCCCCAAGTGGCAACAAAGGACAGAAAGAGAGAGAGGGGGAGCGCT 886

QY 281 GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu 300
 DB 887 GGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACAGCTGCTCCCCCGGAGAG 946
 QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlnProLysGlySerArgGlySerProGly 320
 DB 947 CGTGGCGGCAAGGATCTAAAGGCTCCAGGGCCCAAGGCTCCCGTGGTTCCTCTGGG 1006
 QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
 DB 1007 AAGCCCGGCCCTCAGGGCCCAAGTGGGACCCAGGCCCCCGGCCCCAGGCAAGAG 1066
 QY 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 DB 1067 GGACTCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCAACCGTTGGGAG 1126
 QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 DB 1127 CCTGGGCTGCCCTGGACCTCGGGACTGCCAGGCTTGCCTGGGTACCGATGCCAGGC 1186
 QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 DB 1187 CCCAAGGGCCCCCGGCCCTCTGGCCCATCAGGAGCGGTGTGCCCTGGCCCTGCAG 1246
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 DB 1247 AATGAGCCAAACCCCGCAGCCGAGGACAAATAGCTGCCGCCCTCCTGGAAGAACTTCACA 1306
 QY 421 AspLysCysTyrrTyrrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 DB 1307 GACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGTAGGATGCAAAAGCTTTCTGT 1366
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnTrpIleLys 460
 DB 1367 GAAGACAAGTCTTCACATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAA 1426
 QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 DB 1427 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGAGCGTCAAAAT 1486
 QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrrLysAsnTrpLysAlaGlyGlnPro 500
 DB 1487 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTCAAAAAATTGAAAAAGCTGGACAGCG 1546
 QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrrAlaGly 520
 DB 1547 GATACTGGGTTCATGGCCATGGGCCAGGAGAAAGACTGTGCTGGTTGATTATCTGGG 1606
 QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
 DB 1607 CAGTGGAAACGATTTCCCAATGTGAAGACGCTCAATAACTTCAATTCGTAATAAAGACAGGAG 1666
 QY 541 ThrValLeuSerSerAlaLeu 547
 DB 1667 ACAGTACTGTCTATCTGCATTA 1687

RESULT 5

AAC66903

ID AAC66903 standard; cDNA; 2641 BP.

XX

AC AAC66903;

XX

DT 27-MAR-2001 (first entry)

XX

DE Human EXMAD-14 coding sequence SEQ ID NO: 39.

XX

KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 KW inflammation; reproductive disorder; cardiovascular disorder;
 KW immune disorder; musculoskeletal disorder; developmental disorder;
 KW gastrointestinal disorder; cell proliferation disorder; ss.

OS Homo sapiens.

XX WO200068380-A2.
 XX 16-NOV-2000.
 XX 10-MAY-2000; 2000WO-US12811.
 XX 11-MAY-1999; 99US-0133643.
 XX 23-AUG-1999; 99US-0150409.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 XX Azimzal Y;
 XX WPI: 2001-007395/01.
 XX P-PSDB: AAB27236.
 XX Isolated polynucleotide encoding extracellular matrix or
 XX adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 XX preventing disorders associated with expression of EXMAD such as
 XX proliferative, immune and genetic disorders -
 XX Claim 4; Page 121-122; 129pp; English.
 XX The present invention provides the protein and coding sequences for 25
 XX novel extracellular matrix and adhesion-associated proteins (EXMADs).
 XX These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 XX EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 XX EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 XX EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 XX useful in the prevention and treatment of cancers, cell proliferation,
 XX cardiovascular, reproductive, immune, musculoskeletal, developmental and
 XX gastrointestinal disorders and inflammation.
 XX Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;

Alignment Scores:
 Pred. No.: 1,03e-95 Length: 2641
 Score: 2963.00 Matches: 546
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 99.80% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712A-2 (1-547) x AAC66903 (1-2641)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 Db 297 ATGATTCTCATAATGTGGTTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGCAG 356
 Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 Db 357 CAGAGGAACCTCATCAGCAATCTCGAGCGGTCTGTGGATGACAAAGCCAGGCTATCCAG 416
 Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 417 CGAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGACACG 476
 Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 Db 477 GATTGGCTGAAGGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCCAACAACTCTCGCTTG 536
 Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 537 GCCAAGCCCAACAGCACCTCGAGGATATGAACACCCAGCTCACTCAATTCACAGGT 596
 Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 597 CAGATGGGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGAACCTGAAAGACCTGCAG 656
 Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 1737 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGAAAAAGCTGGACAGCG 1796

Db 657 GACTTACACAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCAACTGAGGAGACGC 716
 Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
 Db 717 TTCAGCTCTTTGAGACGGATATTGTGAACATCATTAACAATATCATGTTACACAGCCAC 776
 Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 777 CACCTGCGGACGCTCACACCAATCTAATGAAGTCAGGACCACTTGCACAGATACCCCTT 836
 Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 837 ACCAAACACACAGATGATCTGACCTCTTGAATAATATCACTCCCTGGCCAACATCCGTTGGAT 896
 Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 897 TCTGTTCTCTCAGGATGCAACAGATTGTAGGTGAGGTGAGGTAGACACTGAAGTAGCC 956
 Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeu 240
 Db 957 AACTTATCAGTGATATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1016
 Qy 241 LysAsnPheThrIleLeuGlnGlnProGlyProArgGlyProArgGlyAspArgGly 260
 Db 1017 AAGAAATTTTACAATACTACAAGGTCCACCGGCCCCAGGGGTCCCAAGAGGTGACAGAGA 1076
 Qy 261 SerGlnGlyProGlyProThrGlyAsnLysGlnLysGlnLysGlnLysGlnLysGlnPro 280
 Db 1077 TCCAGAGGACCCCTTGGCCCACTGGCAACAAAGGACAAAGAGAGAGAGAGGGAGGCT 1136
 Qy 281 GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu 300
 Db 1137 GGACCACTGCGCTCGGGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCCCGGAGAG 1196
 Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
 Db 1197 CGTGGCGCAAGAGATCTAAAGGCTCCAGGGCCCCCAAGGCTCCCGTGGTTCCTCTGGG 1256
 Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGlu 340
 Db 1257 AAGCCCGCCCTCAGGGCCCCAGTGGGAGCCAGGCCCCCCCGGGCCACCAGGCAAGAG 1316
 Qy 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 Db 1317 GGACTCCCGCCCTCAGGGCCCTCTGCTTCCAGGACTTCAGGGCACCCTTGGGGAG 1376
 Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 Db 1377 CCTGGGGTGGCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGC 1436
 Qy 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 Db 1437 CCCAAGGGCCCCCGGCCCTCTCTGGCCCATCAGAGGGGTGGTCCCTGGCCCTGCAG 1496
 Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 Db 1497 AATGAGCCAAACCCCGCACCCGAGGACAATAGTCCCGCCCTCACTGGAAGAACTTACA 1556
 Qy 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 Db 1557 GACAAATCTACTATTTTTTTCAGTTGAGAAAGAAATTTTTTGGAGGATGCAAGCTTTCTGT 1616
 Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
 Db 1617 GAAGCAAGCTCTTCATCTCTTGTTCATAAACACTAGAGAGAACAGCAATGGATAAAA 1676
 Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 Db 1677 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGCCCTCACAGACTCAGAGCGTCAAAAT 1736
 Qy 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
 Db 1737 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGAAAAAGCTGGACAGCG 1796

QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
|||||
Db 1797 GATAACTGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGG 1856

QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
|||||
Db 1857 CAGTGGAGGATTTCAATGTGAAGCGTCAATACTTCATTTGCGAAAAGACAGGGAG 1916

QY 541 ThrValLeuSerSerAlaLeu 547
|||||
Db 1917 ACAGTACTGTCAATCGCATTA 1937

RESULT 6
AAI58842
ID AAI58842 standard; cDNA; 2318 BP.
XX
AC AAI58842;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1045.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB; AAM39686.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1045; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Alignment Scores:

Pred. NO.:	1.63e-95	Length:	2318
Score:	2956.00	Matches:	547
Percent Similarity:	99.82%	Conservative:	0
Best Local Similarity:	99.82%	Mismatches:	0
Query Match:	99.56%	Indels:	1
DB:	22	Gaps:	0

US-09-763-712A-2 (1-547) x AAI58842 (1-2318)

QY	1	MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln	20
Db	47	ATGTATTCTCATTAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAG	106
QY	21	GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln	40
Db	107	CAGAGAACCTCATCCGAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAG	166
QY	41	ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr	60
Db	167	CGAATCAAGAAGACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGACAG	226
QY	61	AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu	80
Db	227	GATTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGACACTGGCTGCCAACAACCTCGCGTTG	286
QY	81	AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly	100
Db	287	GCCAAAGCCCAACACGACACCTCGGAGGATATGAACAGCCACTCACTCATTCACAGGT	346
QY	101	GlnMetGluAsnIleThrThrIleSerGln-AlaAsnGluGlnAsnLeuLysAspLeuGln	120
Db	347	CAGATGGAGAACATCACCACTATCTCTCAAGGCCCAACGAGCAGAACCTGAAAGACCTGCA	406
QY	120	AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGluArg	140
Db	407	GGACTTACACAAAGATGCAGAGAAATAGAACACCCATCAAGTTCAACCAACTCGAGGAAG	466
QY	140	gPheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHi	160
Db	467	CTTCCAGCTCTTTGACAGCGGATATTTGTGAACATCATATTAGCAATATCAGTTACACAGCCA	526
QY	160	shisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrIle	180
Db	527	CCACCTGCGGAGCTGACCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCT	586
QY	180	uThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAs	200
Db	587	TACCAACACACAGATGATCTGACCTCTTGAATTAATACCTGGCCAACTCCGTTTGGG	646
QY	200	pSerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAl	220
Db	647	TTCTGTTTCTCTCAGGATGCAACAAGATTTGATGAGTTCGAGGTTAGACACTGAAGTAGC	706
QY	220	aAsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle	240
Db	707	CAACTTATCAGTGATATATGAAGAAATGAAGCTAGTAGACTCCAAGCATGCTCAGCTCAT	766
QY	240	eLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGln	260
Db	767	CAAGAAATTTTCAATACTACAGGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGG	826
QY	260	ySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPr	280
Db	827	ATCCAGGAGCCCTCGCCCACTGGCAACAGGACAGAAAGAGAGAGAGGGGGAGGCC	886

Qy 280 oGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGly 300
Dbb TGGACCACTGGCCCTGGGGTGGAGAGAGGCCAATTGGACCACTGGTCCGCCCGGAGA 946
Qy 300 uArgGlyGlyGlySerGlySerGlyGlyProGlyGlySerGlySerProGly 320
Dbb CGGTGGCGGCAAGAGATCTAAGAGCTCCAGGGCCCCAAGAGCTCCCGTGGTCCCGCTGG 1006
Qy 320 yLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGly 340
Dbb GAAGCCCGCCCTCAGGCGCCAGTGGGACCCAGGCGCCCGCCGCGCCACAGGCAAGA 1066
Qy 340 uGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGly 360
Dbb GGGACTCCCGCGCCCTCAGGCGCCCTCCCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGA 1126
Qy 360 uProGlyValProGlyProAlaGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Dbb GCTGGGGTGGCTGGACCTCGGGGACTCCAGGCTTGGCTGGGGTACCAAGGATGCCAGG 1186
Qy 380 yProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGly 400
Dbb CCCCAAGGCGCCCGCCGCGCCCTCCCTGGCCCATCAGGAGCGGTGGTGGCCCTGCCA 1246
Qy 400 nAsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPhTh 420
Dbb GAATGAGCAACCCCGCGCCAGGAGCAATGGCTGCCGCGCTCACTGGAAGAACTTCAC 1306
Qy 420 rAspLysCysTrpTyrPheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCys 440
Dbb AGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTG 1366
Qy 440 sGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLeu 460
Dbb TGAACACAAGCTTCCACATCTGTTTTCATAAACAACACTAGAGAGGAACAATGATGATAA 1426
Qy 460 sLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAs 480
Dbb AAAACAGATGTAGGAGAGAGAGCCACTGGATCGGCTCCAGACTCAGAGCTGAGGCTGAAA 1486
Qy 480 nGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnPr 500
Dbb CGAATGGAGTGGTGGTGGATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACACC 1546
Qy 500 oAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Dbb GGATAACTGGGTGATGCCATGGCCAGGAGAGACTGTCTGGGTGATTATGCTGTG 1606
Qy 520 yGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhelLeCysGluLysAspArgGly 540
Dbb GCAGTGGGAACGATTCCCAATGTGAAGACGTCAATAAATTCATTTCGGAAGAGACAGGGA 1666
Qy 540 uThrValLeuSerSerAlaLeu 547
Dbb GACAGTACTGTCATCTGCATTA 1688
RESULT 7
ABQ92072
ID ABQ92072 standard; cDNA: 2929 BP.
XX AC ABQ92072;
XX AC ABQ92072;
DT 04-OCT-2002 (first entry)
XX Human polynucleotide SEQ ID NO 69.
XX Human: cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antitumor; fungicide; antidiabetic; antidiabetic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;

Alzheimer's disease; Parkinson's disease; Huntington's disease;
osteoporosis; severe combined immunodeficiency; SCID; infection;
multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
Homo sapiens.
US2002065394-A1.
30-MAY-2002.
22-DEC-2000; 2000US-0745763.
18-MAR-1998; 98US-0040963.
(JACO/) JACOBS K.
(MCCO/) MCCOY J M.
(LAVA/) LAVALLIE E R.
(COLL/) COLLINS-RACIE L A.
(EVAN/) EVANS C.
(MERB/) MERBERG D.
(TREA/) TREACY M.
(SPAU/) SPAULDING V.

Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
Merberg D, Treacy M, Spaulding V;

WPI; 2002-582343/62.
P-PSDB; ABP61859.

Novel secreted or transmembrane protein and polynucleotide encoding the
protein, useful for diagnosis and treatment of neurological disorders,
cancer, autoimmune diseases, bone disorders and lung or liver fibrosis

Claim 250; Page 231-232; 284pp; English.

The invention relates to human secreted or transmembrane protein (I),
their fragments and is encoded by specific complementary deoxyribonucleic
acid (cDNA) inserts (II), where the protein is substantially free from
other mammalian proteins. (I) are useful for preventing, treating or
ameliorating a medical condition, especially immunological treatment or
prevention of tumors. (I) exhibits activity relating to angiogenesis,
cytokine, cell proliferation, cell differentiation, antiinflammatory,
stem cell growth factor activity and activin or inhibin-related
activities. (I) can be used to manipulate stem cells in culture to give
rise to neuroepithelial cells that can be used to augment or replace
cells damaged by illness, autoimmune disease, accidental damage or
genetic disorders. (I) induces the proliferation of neural cells and
regeneration of nerve and brain tissue and is useful for the treatment of
central and peripheral nervous system diseases and neuropathies, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis. (I) is involved in chemotactic or chemokinetic
activity, regulation of haematopoiesis and is useful for treating myeloid
or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
and for regeneration of bone, cartilage, tendon, ligament and/or nerve
tissue growth and in tissue repair, healing of burns, incisions, ulcers,
for treating osteoporosis, osteoarthritis, bone degenerative disorders or
periodontal disease. (I) is also useful for gut protection or
regeneration and treatment of lung or liver fibrosis, reperfusion injury
in various tissues, various immune deficiencies and disorders including
severe combined immunodeficiency (SCID), bacterial or fungal infections,
autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
such as asthma or other respiratory problems. (II) is useful to express
recombinant protein, as markers for tissues in which the corresponding
protein is preferentially expressed and in gene therapy. The present
sequence is that of a polynucleotide of the invention.

Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:

Pred. No.: 9.74e-95 Length: 2929
Score: 2936.00 Matches: 545

Percent Similarity: 99.63% Conservative: 0
 Best Local Similarity: 99.63% Mismatches: 2
 Query Match: 98.89% Indels: 1
 DB: 24 Gaps: 0

US-09-763-712a-2 (1-547) x ABQ92072 (1-2929)

QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnLeuThrGlnValGln 20
 |||||
 DB 652 ATGATTCATCAATGTGTCATCATGAATC-AACAACCTGNAACCTGACCCAGGTGCAG 710
 |||||

QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 |||||
 DB 711 CAGAGGAACCTCATCAGGAATTCGACGGTCTGTGGATGACACAAGCCAGGCTATCCAG 770
 |||||

QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysAspThr 60
 |||||
 DB 771 CGAATCAAGAACGACATTCCTCAAAATCTGCACGAGTGTTCCTCAAGCCAAAGGACACG 830
 |||||

QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 |||||
 DB 831 GATTGCTGAGGAGAAAGTCGAGACCTTCGACAGCGCTGGCTGCCAACAACCTCGCGTTG 890
 |||||

QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 |||||
 DB 891 GCCAAAGCCAAACAACGACACCTCGGAGGATATGAACAGCCAGCTCAACTCAATTCACAGGT 950
 |||||

QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 |||||
 DB 951 CAGATGGAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAACCTGAAGACCTGCGAG 1010
 |||||

QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 |||||
 DB 1011 GACTTACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCACTGGAGGAACGC 1070
 |||||

QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
 |||||
 DB 1071 TTCACAGCTCTTGAGACGGATATGTGAACATCATTTAGCAATATCATAGTTCACAGCCAC 1130
 |||||

QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 |||||
 DB 1131 CACCTCGGACGCTGACAGCAATCTTAATGAAGTCAGACCACTTGACAGATACCCCT 1190
 |||||

QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 |||||
 DB 1191 ACCAAACACACAGATCATCTGACCTCCTTGAATAATACCTTGCCCAACATCCGTTGGAT 1250
 |||||

QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 |||||
 DB 1251 TCTGTTTCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGTTAGACACTGAAGTAGCC 1310
 |||||

QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 |||||
 DB 1311 AACTTATCAGTATTATGAAGAATGAAGTAGTAGACTCCAAGCATGGTCAGCTCATC 1370
 |||||

QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
 |||||
 DB 1371 AAGAATTTTACATACTACAAGGTCCACGGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGA 1430
 |||||

QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyLysGlyGluPro 280
 |||||
 DB 1431 TCCAGGGAGCCCTCGGCCCACTGGCAACAAGGGACAGAAAGGAGAGGGGGAGCCCT 1490
 |||||

QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlu 300
 |||||
 DB 1491 GGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACCAGCTGTCCTCCCGGAGAG 1550
 |||||

QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
 |||||
 DB 1551 CGTGGCGGAAGAGATCTAAAGCTCCCAAGGGCCCCAAAGGCTCCCGTGGTCCCCCTGGG 1610
 |||||

QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGlu 340
 |||||

DB 1611 AAGCCCGGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCCGGGCCACACGCAAGAG 1670
 QY 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 |||||
 DB 1671 GGACTCCCGGCCCTCAGGGCCCCCTCTGGCTTCCAGGGACTTCAGGGCCCGCTGGGGAG 1730
 |||||

QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 |||||
 DB 1731 CCTGGGCTGCCCTGGACCTCGGGACTGCCAGGCTTCCCTGGGGTACCAGCATGCCAGGC 1790
 |||||

QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 |||||
 DB 1791 CCCAAGGGCCCCCGGCCCTCTCTGGCCCATCAGGAGCGTGTGTCCTGGCCCTGCCAG 1850
 |||||

QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 |||||
 DB 1851 AATGACCAACCCCGCCAGGAGCAATAGTCTGCCCTCCTCTGGAAGAATTTTCTACA 1910
 |||||

QY 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 |||||
 DB 1911 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGT 1970
 |||||

QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
 |||||
 DB 1971 GAAGACAAGCTCTTCACATCTTGTCTTTCATAAACACTAGAGAGCAACGAATGGATAAA 2030
 |||||

QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 |||||
 DB 2031 AACAGATGTAGGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAAT 2090
 |||||

QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
 |||||
 DB 2091 GAATGGAAGTGGCTGGATGGACATCTCCAGACTCAAAAAATTTGAAAGCTGCACAGCCG 2150
 |||||

QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
 |||||
 DB 2151 GATACTGGGGTCATGGCCATGGCCAGGAGAAAGACTGCTGGTGTGATTTATGCTGGG 2210
 |||||

QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
 |||||
 DB 2211 CAGTGAACGATTTCAATGTGAAGACGTCAATACTTCAATTCATTTGCCAAAAAGACAGGAG 2270
 |||||

QY 541 ThrValLeuSerSerAlaLeu 547
 |||||

DB 2271 ACAGTACTGTCATCTGCATTA 2291

RESULT 8
 ID AAV55746 standard; cDNA; 2930 BP.
 AC AAV55746;
 DT 23-MAR-1999 (first entry)
 DE Human secreted protein clone bv227_1 coding sequence.
 KW Secreted protein; human; nutritional supplements; immune stimulant;
 KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
 KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
 KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
 KW cadherin/tumour invasion suppressor; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 67..693
 FT /*tag= a
 XX
 PN WO9855614-A2.
 PD 10-DEC-1998.
 XX 01-JUN-1998; 98WO-US11210.

XX 29-MAY-1998; 98US-0087255.
 PR 04-JUN-1997; 97US-0868696.
 PR 04-JUN-1997; 97US-0868697.
 PR 04-JUN-1997; 97US-0868698.
 PR 04-JUN-1997; 97US-0868898.
 PR 04-JUN-1997; 97US-0868899.
 PR 04-JUN-1997; 97US-0868900.
 PR 04-JUN-1997; 97US-0869191.
 PR 04-JUN-1997; 97US-0869192.
 PR 04-JUN-1997; 97US-0869193.
 PR 04-JUN-1997; 97US-0869194.
 XX (GEMY) GENETICS INST INC.
 XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
 PI McCoy JM, Racie LA, Spaulding V, Treacy M;
 XX WPI; 1999-059912/05.
 DR P-PSDB; AAW73628.
 XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 XX Claim 26; Page 87-88; 127pp; English.
 XX This sequence encodes a human secreted protein of the invention.
 CC This DNA sequence was isolated from a human adult brain cDNA
 CC library, and was designated clone bv227_1. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in
 CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 XX SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;
 Alignment Scores:
 Pred. No.: 9,75e-95 Length: 2930
 Score: 2936.00 Matches: 545
 Percent Similarity: 99.63% Conservative: 0
 Best Local Similarity: 99.63% Mismatches: 2
 Query Match: 98.89% Indels: 1
 DB: 20 Gaps: 0
 US-09-763-712A-2 (1-547) x AAV55746 (1-2930)
 QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 DB 652 ATGTATTCTCAATGTTGGTCATCATGAATC-ACAACCTGAACCTGACCCAGGTGCAG 710
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 DB 711 CAGAGGAACCTCATCACGAATTCGAGCGGTCTGTGGATGACACAAGCCAGGCTATCCAG 770
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 DB 771 CGAATCAAGAACGACTTCAAAATTCGACGAGGTTTCTTCAAGCCCAAGAGACACG 830
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 DB 831 GATTGGCTGAAGGAGAAAGTCAGAGCTTGACAGCGTGGCTGCCAACACTCTCGGTG 890
 QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100

Db 891 GCCAAAGCCAAACAGCACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGT 950
 QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 951 CAGATGGAGAACATCACCACTATCTCTCAAGCCACGACAGCAACCTGAAAGACCTGCAG 1010
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 1011 GACTTACACAAAGATGCAGAGAATAGACAGCCATCAAGTTCAACCACTGGAGGAACGC 1070
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
 Db 1071 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATACAGAGCCAC 1130
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 1131 CACCTGGGCGCTGACCAAGCAATCTAATGAAGTCAGGACCACTTGCACATACCCCT 1190
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 1191 ACCAAACACACAGATGATCTGACCTCCTTGAATAATACCTGCACCAATCCGTTTGGAT 1250
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 1251 TCTGTTCTCTCAGGATGCAACAAGATTTGATGAGGTCAGGTTAGACACTGAAGTAGCC 1310
 QY 221 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 Db 1311 AACTTATCAGTCATTTATGGAGAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1370
 QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGly 260
 Db 1371 AAGAAATTTTACAATACTACAAGGTCCACCGGCCCCAGGGGTCCAAGAGGTGACAGAGGA 1430
 QY 261 SerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyPro 280
 Db 1431 TCCCAAGGACCCCTCGGCCCACTGGCAACAAGGAGGAGAGAGAGAGAGAGAGAGAGCCT 1490
 QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlu 300
 Db 1491 GGACCACTGGCCCTCGGGGTGAGAGAGCCCAATGGACCACTGGTCCCCCGGAGAG 1550
 QY 301 ArgGlyGlyLysSerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
 Db 1551 CGTGGCGGCAAGGATCTAAAGGCTCCCAAGGCCCAAGGCTCCCGTGTTCCCTCGG 1610
 QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
 Db 1611 AAGCCCGGCCCTTCAGGGCCCAAGTGGGACCCAGGCCCCCGGGCCCAAGCAAGAG 1670
 QY 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 Db 1671 GGACTCCCGGCCCTCAGGGCCCTCTGCTTCCAGGGAATTCAGGGACCTGTTGGGAG 1730
 QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 Db 1731 CCTGGGTGCTTGACCTGGGACTGCGAGGCTTGGCTGGGTACCGCATGCCAGC 1790
 QY 381 ProLysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGln 400
 Db 1791 CCCAAGGGCCCCCGGCCCTCTCTGCCCCATCAGGAGCGGTGTGCCCCCTGCGCTGCAG 1850
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 Db 1851 AATGAGCAACCCCGGACCGGAGGACAATAGTGCCTCCCTCCTGGAAGAACTTCACA 1910
 QY 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 Db 1911 GACAAATGCTACTATTTTCACTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTGT 1970
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnTrpIleLys 460
 Db 1971 GAAGACAAGTCTTCATCTTCTTTTCAATAAACACTAGAGAGCAACAGCAATGGATAAA 2030

Qy	461	LysClnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn	480
Db	2031	AAACAGATGTATGGGAGAGAGACCACTGGATCGGCCTCAGAGACTCAGACGCTGAAAT	2090
Qy	481	GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro	500
Db	2091	GAATGGAAGTGGCTGGATGGACATCTCCAGACTACAAAAATGGAAAGCTGGACAGCCG	2150
Qy	501	AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly	520
Db	2151	GATAACTGGGGTCATGCCATGGCCAGGAGAAAGACTGTGCTGGGTTGATTTATGCTGGG	2210
Qy	521	GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluIysAspArgGlu	540
Db	2211	CAGTGGACCATTTCCAAATGTGAAGACGCAATAACTTCATTTGGCAAAAAGACAGGGAG	2270
Qy	541	ThrValLeuSerSerAlaLeu	547
Db	2271	ACAGTACTGTCACTGCATTA	2291
RESULT 9			
AAH43037			
ID	AAH43037 standard; cDNA; 2637 BP.		
AC	AAH43037;		
XX	XX		
XX	XX		
DT	15-OCT-2001 (first entry)		
DE	Nucleotide sequence of a human scavenger receptor.		
XX	XX		
KW	Human; scavenger receptor; SRCL-PI; macrophage; basal immunity;		
KW	arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;		
KW	low density lipoprotein; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
FH	Key		
FT	CDS		
FT	92..2320		
FT	/*tag= a		
FT	/product= "scavenger receptor"		
XX	XX		
PN	W0200159107-Al.		
XX	XX		
PD	16-AUG-2001.		
XX	XX		
PF	08-FEB-2001; 2001WO-JP00874.		
XX	XX		
PR	14-FEB-2000; 2000JP-0035155.		
PR	10-OCT-2000; 2000JP-0309068.		
XX	XX		
PA	(FUSO) FUSO PHARM IND LTD.		
XX	XX		
PI	Wakamiya N;		
XX	XX		
DR	WPI: 2001-497076/54.		
DR	P-PSDB; AAG63347.		
XX	XX		
PT	New Scavenger receptor proteins SRCL-PI with collectin-like structure,		
PT	useful for treatment and diagnosis of diseases associated with oxidized		
PT	low-density lipoprotein accumulation -		
XX	XX		
PS	Claim 6; Page 88-93; 118pp; Japanese.		
XX	XX		
CC	The present sequence encodes a human scavenger receptor, designated		
CC	SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They		
CC	are useful in clarifying the functions of macrophages and basal		
CC	immunity. They are also useful in the treatment, prevention, diagnosis		
CC	and investigation of diseases such as arteriosclerosis, diabetic		
CC	complications, bacterial infection and restenosis following angioplasty,		
CC	which are associated with accumulation of oxidized low density		
CC	lipoprotein and the binding of advanced glycation end-products into		
CC	cells.		

XX	SQ	Sequence	2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;
Alignment Scores:			
Pred. No.:	8,44e-89	Length:	2637
Score:	2765.00	Matches:	502
Percent Similarity:	96.71%	Conservative:	27
Best Local Similarity:	91.77%	Mismatches:	18
Query Match:	93.13%	Indels:	0
DB:	22	Gaps:	0
US-09-763-712A-2 (1-547) x AAH43037 (1-2637)			
Qy	1	MetTyrSerHisAsnValVallIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln	20
Db	677	ATGTATTCTCAGACGGTGTTATCATCAACCTCAACAACCTGAACCTAACCCAGGTTTCAG	736
Qy	21	GlnArgAsnLeuLethrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln	40
Db	737	CAGAGGAACCTTATCTCAATCTGCAGCAGTCTGTGGATGACAAAGCGCTGCCATCCAG	796
Qy	41	ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr	60
Db	797	CGAATTAAAGAATGATTCCAAAATCTGCAGCAGGTTTTCCITTCAGACCAGGAAGACACC	856
Qy	61	AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu	80
Db	857	GATTGGCTAAAGGAAAAAGTACAGAGCTTGCAGACATTTGGCTGCCAACCAACTCTGCCCTG	916
Qy	81	AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly	100
Db	917	GCCAAGCCAAACATGACACCTTAGAGGATATGAATAGCCACGCTCAGCTCATTCACAGGT	976
Qy	101	GlnMetGluAsnLethrThrlleSerGlnAlaAsnGluInAsnLeuLysAspLeuGln	120
Db	977	CAGATGACAAACATTACCACTATCTCACAGGCCAACGAGCAGAGCGCTGAAAAGCCTTCAG	1036
Qy	121	AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg	140
Db	1037	GACTTACCAAGGATACAGAAATAGACAGCTGTCAAGTTCAAGCCAACTTTGAGGAACGC	1096
Qy	141	PheGlnLeuPheGluThrAspIleValAsnIlelleSerAsnIleSerTyrThrAlaHis	160
Db	1097	TTCAGGCTCTTTGAGACAGATAATTGTGAACATCATTAGCAACATCAGCTACACAGCCAT	1156
Qy	161	HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu	180
Db	1157	CACCTGAGGACACTGACCAACATCTGAATGATGTAGGACCATGACACAGACACCTTG	1216
Qy	181	ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp	200
Db	1217	ACCAGACACACGGATGACCTGACCTCCTTGAATACACACATAGTACACATCCGCTTGGAT	1276
Qy	201	SerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAla	220
Db	1277	TCTATTTCTCTCAGGATGCAGCAAGACATGATGAGTCAAAGTTAGACACTGAAGTGCC	1336
Qy	221	AsnLeuSerVallIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle	240
Db	1337	AACATTATCAGTTGTTAGGAAGATATGAACATGGTTGACTCCAAAGCAGCGTCAGCTCATC	1396
Qy	241	LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly	260
Db	1397	AGAGACTTACCATTCTACAAGGTCTCTCTGCCCCAGAGGTTCCAAAAGGTGCACAGAGA	1456
Qy	261	SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro	280
Db	1457	TCTCAGGACCACTGGTCCAACTGSCAACAAAGGCACAGAAAGGAGAGAAAGGAGAGCCT	1516
Qy	281	GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu	300
Db	1517	GGTCCACTGGCCCTCGGGGTGAGAGGGGCAAAATTTGACCATGCGGCCCTCTCTGGAGAG	1576

```

Qy 301 ArgGlyGlyLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1577 CGTGCAGCAAAAGGATCAAAAGGCTCACAGGTCCTCAAGGATCTCGTGGTCCCGAGG 1636
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProGlyLysGlu 340
Db 1637 AAGCCTGGCCCTCAAGGACCTAGTGGGACCCAGGACCAAGGTCACAGGCAAGGAT 1696
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1697 GGACTCCCTGGCCCTCAGGCGCCCTCCTGGCTCCAGGACCTACAGGCACTGTGGTGAG 1756
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1757 CCTGAGTACCTGGACCTCGGGGGTGGCAGGCTTGCAGGGTGCAGGCACTGGCTGGG 1816
Qy 381 ProLysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGln 400
Db 1817 CCTAAGGACCACTGGCCCTCAGGCGCCCTCAGGAGCAATGGCCCACTGGCTTGCAG 1876
Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db 1877 ATGAACCAACCCAGCATCAGAGTCAACGGATGTCGCCCTCATCGGAAGACTTCACA 1936
Qy 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1937 GATAAATGCTACTATTTTCATTGGAAGAAATTTTGAAGATGCTAAGCTTTCTCTGT 1996
Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db 1997 GAAGACAAATCTCCCATCTCGTTTTCATAAATCAAGAGAAGAACAGCAATGGATAAA 2056
Qy 461 LysGlnMetValGlyArgGlySerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2057 AGCATACCGTGGGAGAGAACCATTTGGATCGCCCTCAGAGTCAAGACAGGAAAGC 2116
Qy 481 GluTrpLysTyrLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
Db 2117 GAATGGAAGTGGCTAGACGGGTCACTGTGTATTACAAAACTGGAAGCTGGACAACCA 2176
Qy 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2177 GATAACTGGGCGAGTGGCCATGGCCAGGAGAGACTGTGTGGCTGATTTACGACGGA 2236
Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2237 CAGTGAATGACTTCCAGTGTGATGAATCAATACTTCACTTTGTGAGAGAGGAGGAG 2296
Qy 541 ThrValLeuSerSerAlaLeu 547
Db 2297 GCAGTACCATCATCATATTA 2317

RESULT 10
ABL95574
ID ABL95574 standard; cDNA; 2005 BP.
XX
AC ABL95574;
XX
XX
DT 19-JUL-2002 (first entry)
XX
DE Human anglogenesis related cDNA PRO7223 SEQ ID NO: 27.
XX
KW Human; anglogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;
KW antiarteriosclerotic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200208284-A2.
XX
XX 31-JAN-2002.
PD
XX

```

```

PF 09-JUL-2001; 2001WO-US21735.
XX
XX 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001US-0506666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 30-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX
XX (GETH ) GENENTECH INC.
PA (BAKE) BAKER K P.
PA (FERR) FERRARA N.
PA (GERB) GERBER H.
PA (GERR) GERRITSEN M E.
PA (GODD) GODDARD A.
PA (GODO) GODOWSKI P J.
PA (GURN) GURNEY A L.
PA (HILL) HILLAN K J.
PA (MARS) MARSTERS S A.
PA (PANJ) PAN J.
PA (PAON) PAONI N F.
PA (STEP) STEPHAN J F.
PA (WATA) WATANABE C K.
PA (WILL) WILLIAMS P M.
PA (WOOD) WOOD W I.
XX
XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Maisters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WL, Ye W;
XX
XX WPI; 2002-171999/22.
DR P-PSDB; ABB95436.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
XX Claim 1; Fig 27; 567pp; English.
PS
XX The present invention provides the protein and coding sequences of human
CC PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac

```

CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
CC healing. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Alignment Scores: 3,35e-86 Length: 2005
Pred. No.: 2688.00 Matches: 495
Score: 2688.00 Mismatches: 2
Percent Similarity: 99.60%
Best Local Similarity: 99.20%
Query Match: 90.54%
DB: 24
 24
 0
 0

US-09-763-712A-2 (1-547) x ABL95574 (1-2005)

QY 49 LeuGlnGlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGln 68
DB 8 GTCGGCAGGTTTCTTCAAGCCAAAGAGGACACGGATTGGCTGAAGGAGAAAGTGCAG 67
QY 69 SerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAsnAspThrLeu 88
DB 68 AGCTTCGACACACTGGCTGCCAACAACTCTGCTTGGCCAAAGCCCAACACACACCCCTG 127
QY 89 GluAspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIle 108
DB 128 GAGGATATGAACAGCCAGCTCAACTCATTACAGGTCAGATGGAGAATCACCACATC 187
QY 109 SerGlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsn 128
DB 188 TCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAAT 247
QY 129 ArgThrAlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIle 148
DB 248 AGAACAGCCATCAAGTTCAACCACTGGAGAACGCTTCCAGCTCTTTGAGCGGATATT 307
QY 149 ValAsnIleLeuSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsn 168
DB 308 GTGAACATCATAGCAATATCAGTTACACAGCCACCACTGCGGACGCTGACCAACAT 367
QY 169 LeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThr 188
DB 368 CTAAATGAAGTCAGGACCACTTGCACAGATACCTT-ACCATAACACACAGATCTGACC 426
QY 189 SerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGln 208
DB 427 TCCTTGAATATACCTGGCCACACATCCGTTGGATTCTGTTCTCTCAGGATGCCAA 486
QY 209 AspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGlu 228
DB 487 GATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGAAGA 546
QY 229 MetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGly 248
DB 547 ATGAAGCTAGTACCTCAAGCATGGTCAGCTCATCAAGATTTTACAATACTACAAGT 606
QY 249 ProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThr 268
DB 607 CCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCAGGGACCCCTGGCCCAACT 666
QY 269 GlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGlu 288
DB 667 GCACAACAGGGACAGAAAGGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAG 726
QY 289 ArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGly 308
DB 727 AGAGGCCCAATTTGGACACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAAGC 786
QY 309 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 328
DB 787 TCCACAGGGCCCCAAAGGCTCCCGTGGTCCCTTGGAAAGCCCGGCTCAGGGCCCCAGT 846

QY 329 GlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyPro 348
DB 847 GGGGACCCAGGCCCCCGGGCCACCAGGCAAGAGAGGACTCCCGGCCCTCAGGGCCCT 906
QY 349 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGly 368
DB 907 CCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGAGCGCTGGGGTGCCTGACCTCGGGA 966
QY 369 LeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProPro 388
DB 967 CTGCCAGGCTTGCTGGGTACCCAGGCACTGCCAGGGCCCCAAGGGCCCCCGGCCCTCT 1026
QY 389 GlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu 408
DB 1027 GGCCCATCAGGAGCGTGGTGGCCCTGCCCTGCAGATGAGCCACCCCGGCACGGAG 1086
QY 409 AspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerVal 428
DB 1087 GACAATAGCTGCCGCTCCTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTT 1146
QY 429 GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuVal 448
DB 1147 GAGAAAGAAATTTTGAGGATGCAAGCTTCTGTGAAGACAAGTCTTCACATCTGTT 1206
QY 449 PheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer 468
DB 1207 TTCATAACACATAGAGAGGAACAGCAATGGATAAAAAACAGATGCTAGGAGAGAGAC 1266
QY 469 HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr 488
DB 1267 CACTGTGATCGGCTCACAGACTCAGACGCTGAAATGAAATGAAATGAAATGAAATG 1326
QY 489 SerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly 508
DB 1327 TCTCCAGACTACAAATTTGAAAGCTGGACCCGGTAACCTGGGGTCTATGCCCATGG 1386
QY 509 ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
DB 1387 CCAGGAGAAGACTGTCTGGTGTGATTATGCTGGCAGTGGAACCATTTCCAAATGTGA 1446
QY 529 AspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
DB 1447 GACGTCAATAACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCTGCA 1503
RESULT 11
ABL88085
ID ABL88085 standard; cDNA; 2005 BP.
XX ABL88085;
XX
DT 16-MAY-2002 (first entry)
XX
DE Human PRO7223 cDNA sequence SEQ ID NO:27.
XX
KW Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;
KW vulnerable; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW anglogenetic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial restenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping; gene; ss.
XX Homo sapiens.
XX
PN WO200200690-A2.
XX
PD 03-JAN-2002.
XX
PF 20-JUN-2001; 2001WO-US19692.
XX
PR 23-JUN-2000; 2000US-213637P.

DB:	24	Gaps:	0
US-09-763-712A-2 (1-547) x ABL88085 (1-2005)			
QY	49	LeuGlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGln	68
Db	8	GTCCGCAGGCTTTCTTCAAGCCAAAGAAGACACGGATTGGCTGAAGGAGAAAGTCAG	67
QY	69	SerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAsnAspThrLeu	88
Db	68	AGCTTGCAGACACTGGCTGCCAACAACTCTGCCGTGGCCAAAGCCAAACGACACCCCTG	127
QY	89	GluAspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIle	108
Db	128	GAGGATATGAACGCCAGCTCAACTCATTCACAGGTCAGATGGAGAACATCACCACATATC	187
QY	109	SerGlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsn	128
Db	188	TCTCAAGCCAAACGACGAGAACCTTGAAGACCTTCAGGACTTACACAAAGATGCAGAGAAT	247
QY	129	ArgThrAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGlnThrAspIle	148
Db	248	AGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCAGGCTCTTTGAGACGGATATT	307
QY	149	ValAsnIleLysSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsn	168
Db	308	GTGAACATCATTTAGCAATATCACTTACACAGCCACCACCTGCGGACGCTGACCACCAAT	367
QY	169	LeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThr	188
Db	368	CTAAATGAAGTCAAGACCACTTCCACAGATACCTT-ACCAAAACACACAGATGATCTGACC	426
QY	189	SerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGln	208
Db	427	TCCTTGATTAATACCTTGGCCAACTCCGTTGGATCTGTCTCTCAGGATGCAACAA	486
QY	209	AspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGlu	228
Db	487	GATTTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTATTATGGAAGAA	546
QY	229	MetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGly	248
Db	547	ATGAAGCTAGTAGACTCCAAGCATGTGTAGCTCATCAAGAAATTTTACAATACTACAAAGT	606
QY	249	ProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThr	268
Db	607	CCACGGGGCCCAAGGGTCCACAGAGTGCACAGAGATCCACAGGACCCCTCGGCCCACT	666
QY	269	GlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGlu	288
Db	667	GGCAACAAGGCACAGAAAGAGAGAGAGGGGGAGCCTGGACCCTGGCCCTGGCGGTGAG	726
QY	289	ArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGly	308
Db	727	AGAGGGCCCAATTGGACAGCTGTGTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGC	786
QY	309	SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer	328
Db	787	TCCAGGGCCCAAGGCTCCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCCAGT	846
QY	329	GlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyPro	348
Db	847	GGGGACCCAGGCCCCCGGGCCACCAGGCAAGAGAGGACTTCCCGCGCCTCAGGGCCCT	906
QY	349	ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGly	368
Db	907	CCGTGGCTTCCAGGACTTCAAGGCACCGTTGGGAGCCCTGGGTGGCTCGGGGA	966
QY	369	LeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProPro	388
Db	967	CTGCCAGGCTTGCTGGGTATCCAGGCATGCCAGGGCCCCAAGGGCCCCCGGCCCTCTCT	1021
QY	389	GlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu	408

Pred. No.:	3,35e-86	Length:	2005
Score:	2688.00	Matches:	495
Percent Similarity:	99.60%	Conservative:	2
Best Local Similarity:	99.20%	Mismatches:	2
Query Match:	90.54%	Indels:	1

Db 1027 GGCCCATCAGAGCGGTGGCCCTGGCCCTCAGAAATGAGCAACCCCGGACCGGAG 1086
 QY 409 AspAsnGlyCysProHisTrpLysAsnPhenThrAspLysCysTrpLysPheSerVal 428
 Db 1087 GACAATAGCTGCCGCTCACTGGAAGACTTCACAGACAAATGCTACTATTTTTCAGTT 1146
 QY 429 GluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuVal 448
 Db 1147 GAGAAAGAAATTTTGGAGTGCAGAAAGCTTTCTGTCAAGACAAAGCTTCCATCTTGT 1206
 QY 449 PheLeuAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer 468
 Db 1207 TTCATAAACACTAGAGAGGAACAGCATGATGATAAAAAACAGATGTTAGGAGAGAGAGC 1266
 QY 469 HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLysPheSerGlyThr 488
 Db 1267 CATGTGATCGGCTCACAGACTCAGAGCGTGAATGAAATGGAAGTGGCTGGATGGGACA 1326
 QY 489 SerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly 508
 Db 1327 TCTCCAGACTACAAAATTTGAAAGCTGGAGACGCCGATAACTGGGTCTATGCCATGGG 1386
 QY 509 ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
 Db 1387 CCAGGAGAAAGACTGTGCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTCCAAATGTGA 1446
 QY 529 AspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
 Db 1447 GAGTCNTAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTATTA 1503
 RESULT 12
 AAH43054
 ID AAH43054 standard; DNA: 2256 BP.
 XX
 AC AAH43054;
 XX
 XX 15-OCT-2001 (first entry)
 DT
 XX Nucleotide sequence of a human scavenger receptor.
 DE
 XX Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.
 XX
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 74..1936
 FT /*tag= a
 FT /product= "scavenger receptor"
 XX
 XX WO200159107-A1.
 PD 16-AUG-2001.
 XX
 XX 08-FEB-2001; 2001WO-JP00874.
 XX
 XX 14-FEB-2000; 2000JP-0035155.
 PR 10-OCT-2000; 2000JP-0309068.
 PR
 XX (FUSO) FUSO PHARM IND LTD.
 PA
 XX Wakamiya N;
 PI
 XX WPI: 2001-497076/54.
 DR P-PSDB; AAG63350.
 DR
 XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 PT useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation
 XX
 PS Claim 4; Page 105-109; 118pp; Japanese.

XX The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.
 XX
 SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;
 Alignment Scores: Length: 2256
 Pred. No.: 1,15e-68 Matches: 423
 Score: 2187.00 Conservative: 0
 Percent Similarity: 77.33% Mismatches: 0
 Best Local Similarity: 77.33% Indels: 124
 Query Match: 73.66% Gaps: 1
 DB: 22
 US-09-763-712A-2 (1-547) x AAH43054 (1-2256)
 QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuThrGlnValGln 20
 Db 659 ATGTATTCTCATATGTGGTCAATCATCAACCAACCTGAACCTGACCCAGGTGCAG 718
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
 Db 719 CAGAGGAACCTCATACGAAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAG 778
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 779 CGAATCAAGAACGACATTCATAATCTGCAGCAGGTTTTTCTCAAGCCAAAGAGGACAG 838
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 Db 839 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGCAGCTGGCTGCCAACAACTCTGCGCTTG 898
 QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 899 GCCAAAGCCCAACCAACGACACCTGGAGGATATGAACAGCCGCTCACTCATTCACAGGT 958
 QY 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 959 CAGATGGAGAACATCACCACCTATCTCTCAAGCCCAAGCAGCAGAACCTGGAAGACCTGCAG 1018
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 1019 GACTTACACAAAGATGCAGAGAAATAGAACCCATCAAGTTCAACCAACTGGAGGACGC 1078
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
 Db 1079 TTCAGCTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTACAGGCCAC 1138
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 1139 CACCTGGGACGCTGACCAACATCTAAATGAAGTCAAGGACCACTTGCACAGATACCTT 1198
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 1199 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTTGGCCCAACATCCCTTTGGAT 1258
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuValAla 220
 Db 1259 TCTGTTTCTCTCAGGATGCACCAAGATTTGATAGGTCGAGGTATGACACTGAAGTAGCC 1318
 QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 Db 1319 AACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGTGTCAGTCATC 1378
 QY 241 LysAsnPheThrIleLeuGlnGlnProGlyProArgGlyProArgGlyAspArgGly 260

QY 282 ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg 301
 Db 277 CCACCTGGCCCTCGCGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGGAGAGCGT 336
 QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
 Db 337 GCGCGCAAGAGTCTAAAGCTCCAGGGGCCCAAGAGGCTCCGCTGGTTCCTCCCTGGGAG 396
 QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGly 341
 Db 397 CCGGCGCCCTCAGGGCTCCAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGGA 456
 QY 342 LeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 361
 Db 457 CTCGCGGCGCCCTCAGGGCCCTCTGGCTCCAGGGACATTCAGGGCACCGTTGGGGAGCCT 516
 QY 362 GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro 381
 Db 517 GGGGTGCTTGGACCTCGGGGACTGCGAGCTTGGCTGGGGTACCAGGCATGCGAGGCCCC 576
 QY 382 LysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsn 401
 Db 577 AAGGGCCCCCGGCGCTCTGGCCATCAGAGGCGGTGGTCCCTGCGCTGCAGAAAT 636
 QY 402 GluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAsp 421
 Db 637 GAGCAACCCAGCAGCCGAGGACAAATGGCTGCCCGCTCACTGGGAAGAACTTCACAGAC 696
 QY 422 LysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGlu 441
 Db 697 AAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGAGCTTTCTGTGAA 756
 QY 442 AspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLys 461
 Db 757 GACAGCTCTTCACATCTGTTTTCATTAACACTAGAGGAGACACAAATGGATTAATAAAA 816
 QY 462 GlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGlu 481
 Db 817 CAGATGGTAGGAGAGAGAGGACCTGGATGGCCCTCACAGACTCAGAGCGTGAAAAATGAA 876
 QY 482 TrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAsp 501
 Db 877 TGGAGTGGCTGGATGGGACATCTCCAGACTACAAAAATGGAAAGCTGGAGCGCGGAT 936
 QY 502 AsnTrpGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGln 521
 Db 937 AACTGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTCATTTATCTGGGCAG 996
 QY 522 TrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThr 541
 Db 997 TGGACGATTTCCAAATGTGAAGACGTCAATAACTTCAATTTGGAAAAAGACAGGGAGACA 1056
 QY 542 ValLeuSer 544
 Db 1057 GTCTTCAGC 1065
 RESULT 14
 AAX27858
 ID AAX27858 standard; DNA; 3685 BP.
 XX
 AC AAX27858;
 XX
 DT 02-JUN-1999 (first entry)
 DE Human CSR3 protein coding sequence.
 XX
 KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
 KW scavenger receptor protein; intracellular stress; arteriosclerosis;
 KW diabetic circulatory obstruction; microbial infection; ss.
 OS Homo sapiens.
 XX
 PN WO9909159-A1.

XX 25-FEB-1999.
 PD
 XX 12-AUG-1998; 98WO-JP03602.
 PF
 XX 30-JUL-1998; 98JP-0230121.
 PR
 PR 13-AUG-1997; 97JP-0233396.
 XX
 PA (NLSB) JAPAN TOBACCO INC.
 XX
 PI Nakamura Y, Tokino T;
 DR WPI; 1999-181032/15.
 DR P-PSDB; AAY00994.
 XX
 PT Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress
 XX
 PS Claim 6; Page 142-150; 175pp; Japanese.
 XX
 CC This sequence encodes the human cellular stress response 3 (CSR3) protein
 CC of the invention. The CSR proteins are macrophage scavenger receptor
 CC proteins. The CSR proteins can be used in the treatment, gene therapy
 CC and diagnosis of diseases in which intracellular stress is important,
 CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
 CC infection. Expression of the proteins is induced in vivo in response to
 CC intracellular stress, and inhibits cell death as a result of such stress.
 XX
 SQ Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other;
 Alignment Scores:
 Pred. NO.: 7.76e-19 Length: 3685
 Score: 765.00 Matches: 154
 Percent Similarity: 60.41% Conservative: 81
 Best Local Similarity: 39.59% Mismatches: 134
 Query Match: 25.77% Indels: 20
 DB: 20 Gaps: 4
 US-09-763-712A-2 (1-547) x AAX27858 (1-3685)
 QY 11 LeuAsnAsnLeuAsnLeuThrGlnValGlnArgAsnLeuIleThrAsnLeuGlnArg 30
 Db 808 GTGCACCAAGATCAATTCACCGTGGGAGAGACTTCGAGTGGATCCAGGGATCCAGCGG 867
 QY 31 SerValAspAspThrSerGlnAlaIleGlnArgIleLysAsnAspPheGlnAsnLeuGln 50
 Db 868 ANGACAGCAGGAGAGACCTGACCTCCAGAGAGATTGTCCACCGACTGGCAGAACTACACA 927
 QY 51 GlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGlnSerLeu 70
 Db 928 CGGCTCTTCAGCGGCTGCGCACCACTCCACCAAGACTGGAGAGCGGCTCAAGAACATC 987
 QY 71 Gln---ThrLeuAlaAlaAsnSerAlaLeuAlaLysAlaAsnAspThrLeuGlu 89
 Db 988 CAGGCCACCCCTGGGGCTCTCCAGCGCATCAGCCAG---AACTCAGAGAGCATGCAC 1044
 QY 90 AspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGlnAsnIleThrThrIleSer 109
 Db 1045 GACCTGGTACTCCAGTCCAGTCCAGTGGCTGGAGTGCAGTGGATAACATCTCGTCTTCTG 1104
 QY 110 GlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArg 129
 Db 1105 GATGACCAAGAGAGAACATGCATGATCTTCAGTACCATCCACTACCCAGCCAGAACCGC 1164
 QY 130 ThrAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleVal 149
 Db 1165 ACTGTGGAGAGGTTTGAGTCTCTGGAAGGAGCCATCGGCTTCTCAGCAGAGATTGAATGGC 1224
 QY 150 AsnIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeu 169
 Db 1225 ACCATCTTCACCAACATCAATGCCACCGACCAACACCGTGCACAGCATGCTCAAGTACCTG 1284
 QY 170 AsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSer 189

```
Db 1285 GATGAGCTGCGGCTCTCTCCGACCTGGGCTTCCACACCCATCGCGAGGAGCTCTACTAC 1344
Qy 190 LeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAsp 209
Db 1345 CTGACAAAGTCTGCTCCATCATGCTGGGCACACAGACCTCTCGGAGGCGCTTCAGC 1404
Qy 210 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 229
Db 1405 CTGCTCAGTCCGCGCTGGACCTCAACGTCGCGAACCCTCTCCATGATGCTGGAGGAGATG 1464
Qy 230 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro 249
Db 1465 AAGGAGTGGACACACAGCATGGAGAAATCTTCGCAATGTCCACCATCTCAGAGGTGCC 1524
Qy 250 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGly 269
Db 1525 CCGCGCCCTCCAGGACCAAGAGATCAAGAGGATATGGGGCTGNAAGGGCTGTGGC 1584
Qy 270 AsnLysGlyGlnLysGlyGlu-----Lys 277
Db 1585 GGCAGAGGCCCGAAAGGAGACCCGCGCATCTTGGGCCCTGGGACCCCGAGGTCCTCAG 1644
Qy 278 GlyGluProGlyProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro 297
Db 1645 GGGCAACCTGGAGAGCGCGGCTGTGGAGAAAGGGCCCTGTGGCCCTCGAGGGTTC 1704
Qy 298 ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly 317
Db 1705 CCAGGCTCAAGGCTCAAGGCGAGCTTGGAACTGGAGGCCGAGG-----Lys 1752
Qy 318 SerProGlyLysProGlyProGlyProGlyProSerGlyAspProGlyProProGlyProPro 337
Db 1753 -----GGACAGCCAGGCCCAAGAGGGACATAGGGCCCGCCAGGCGCAGAGGGCCCGG 1806
Qy 338 GlyLysGluGlyLeuProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThr 357
Db 1807 GGGTCTCCAGGGCCCTCAGGGCTCAGGGAAACCGGAAATTCAGAGGAAGACAGGGTCA 1866
Qy 358 ValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 377
Db 1867 CCAGGCCAGCGGGGCCCTAGGGCTAGAGGTGAACAGGATCCAGGGTCCCGCTGTGT 1926
Qy 378 MetProGlyProLysGlyProGlyProGly 386
Db 1927 CTCGCGGGGCTCCAGGTCACACGGA 1953

RESULT 15
AAAX27856
ID AAAX27856 standard; DNA; 3810 BP.
XX
AC AAX27856;
XX
DT 02-JUN-1999 (first entry)
XX
DE Human CSR1 protein coding sequence.
XX
KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX
OS Homo sapiens.
XX
PN W09509159-A1.
XX
PD 25-FEB-1999.
XX
PF 12-AUG-1998; 98WO-JP03602.
XX
PR 30-JUL-1998; 98JP-0230121.
XX
PR 13-AUG-1997; 97JP-0233396.
XX
PA (NIBS ) JAPAN TOBACCO INC.
```

```
XX Nakamura Y, Tokino T;
XX WPI: 1999-181032/15.
DR P-PSDB; AAY00992.
XX
PT Scavenger receptor proteins.- for treatment and diagnosis of
PT disorders involving cell stress
XX
PS Claim 4; Page 119-127; 175pp; Japanese.
XX
CC This sequence encodes the human cellular stress response 1 (CSR1) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
SQ Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;

Alignment Scores:
Pred. No.: 7,96e-19 Length: 3810
Score: 765.00 Matches: 154
Percent Similarity: 60.41% Conservative: 81
Best Local Similarity: 39.59% Mismatches: 134
Query Match: 25.77% Indels: 20
DB: 20 Gaps: 4

US-09-763-712A-2 (1-547) x AAX27856 (1-3810)
Qy 11 LeuAsnAsnLeuAsnLeuThrGlnValGlnGlnArgAsnLeuIleThrAsnLeuGlnArg 30
Db 933 GTGCACCAGCTCAACTTCCCGTGGGGCAGACTTCCGAGTGGATCCACGGATCCAGCG 992
Qy 31 SerValAspAspThrSerGlnAlaIleGlnArgIleLysAsnAspPheGlnAsnLeuGln 50
Db 993 AAGACAGACGAGGAGACCTGACCTCCAGAGAGATTGTCCACGACTGGCAGAACTACACA 1052
Qy 51 GlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGlnLysValGlnSerLeu 70
Db 1053 CGGCTCTTCGCGGCGCTGGCGCACCTCCACCAAGACTGGAGAGCGGCTCAGAAACATC 1112
Qy 71 Gln---ThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAspThrLeuGlu 89
Db 1113 CAGGCCACCTGGGGGCTCTCCACAGCGCATCAGCGAG---AACTCAGAGACCATGCAC 1169
Qy 90 AspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIleSer 109
Db 1170 GACCTGGTACTCCAGGTCATGGGCTTGCAGCTGCAGCTGGATGAACATCTCGTCTCTG 1229
Qy 110 GlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArg 129
Db 1230 GATGACCAAGAGAGAACATGATGATTCAGTACCATACCCACTACGCCCAAGAACCGC 1289
Qy 130 ThrAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleVal 149
Db 1290 ACTGTGGAGAGGTTTCACTCTCTGGAGAGCGCATGGCTTCTCAGGATTTGAATGGC 1349
Qy 150 AsnIleIleSerAsnIleSerTyThrAlaHisHisLeuArgThrLeuSerAsnLeu 169
Db 1350 ACCATCTTCCCAACATCAATGCCACCGACACACCGTGCACAGCATGCTCAAGTACCTG 1409
Qy 170 AsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSer 189
Db 1410 GATGAGTGGGCTCTCTCCGACGCTGGGCTTCCACACCCATCGCGAGGAGCTCTACTAC 1469
Qy 190 LeuAsnAspThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAsp 209
Db 1470 CTGAACAAGTCTGTCTCCATCATGCTGGGCACACAGACCTGCTCGGGAGCGCTTCAGC 1529
Qy 210 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 229
Db 1584 GATGAGTGGGCTCTCTCCGACGCTGGGCTTCCACACCCATCGCGAGGAGCTCTACTAC 1644
```

Db	1530	CTGCTCAGTGGCCGGCTGGACCTCAACGTCGGAACCTCTCCATGATCGTGAGGAGATG	1589
QY	230	LysLeuValAspSerLysHisGlyLeuLeuLeuLysAsnPheThrIleLeuGlnGlyPro	249
Db	1590	AAGCGAGTGGACACACAGCATGGAGAAATCTCTCCATGTCACCATCTACGAGGTGCC	1649
QY	250	ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGly	269
Db	1650	CCCGGCCCTCCAGGACCAAGAGGATTCAAAGGAGATATGGCGGTGAAGGGCCTGTTGGC	1709
QY	270	AsnLysGlyClnLysGlyGlu-----Lys	277
Db	1710	GGCAGAGGCCCCAAAAGAGACCCCGGCATCTTTGGCCGCCCTGGACACCCAGGGTCTCTCAG	1769
QY	278	GlyGluProGlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro	297
Db	1770	GGCACACTGGAGAGGCCGGGCTGTGGAGAAAGGGCCCTGTGGCCCTCGAGGGGTTG	1829
QY	298	ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly	317
Db	1830	CCAGGCTCAAAGGCTCAAAGGCGAGCTTTGGAACCTGGAGGGCCGAGA-----	1877
QY	318	SerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProPro	337
Db	1878	-----GGACAGCCAGCGCCCAAAGGGACATAGGGCCCCCCAGGGCCAGAGGGCCCCG	1931
QY	338	GlyLysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThr	357
Db	1932	GGGTCTCCAGGGCCCTCAGGGGCTCAGGGAACCGGAATTGGAGGAAGACAGGGGTCA	1991
QY	358	ValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly	377
Db	1992	CCAGGCCAGCGGGGCCCATGGGGCTTAGGGTGAACAGAGGATCCAGGTCGCCCTTGGT	2051
QY	378	MetProGlyProLysGlyProProGly	386
Db	2052	CTCCGGGGCCTCCAGGTCCACACAGGA	2078

Search completed: March 21, 2003, 08:37:40
Job time : 318.244 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:26:21 ; Search time 58.1137 Seconds
(without alignments)
2886.622 Million cell updates/sec

Title: US-09-763-712A-2

Perfect score: 2969

Sequence: 1 MYSHVYIMNLNLITQVQ.....EDVNNFICKDRTVLSSAL 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09763712/runat_14032003_100951_18187/app_query_fasta_1.1877
-DB=IssuedPatents_NA -QMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -CGN_1_108 @runat_14032003_100951_18187 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	459.5	15.5	810	1	US-08-642-255-60
2	452.5	15.2	3181	1	US-08-655-086-1
3	448.5	15.1	756	1	US-08-642-255-50
4	436.5	14.7	1608	4	US-09-029-348-19
5	428	14.4	4031	1	US-08-159-784-1
6	424.5	14.3	1572	4	US-09-297-269-39
7	419	14.1	2543	1	US-08-555-669-11
8	419	14.1	2543	3	US-09-073-663-11
9	412.5	13.9	432	1	US-08-642-255-48
10	409	13.8	3394	1	US-08-159-784-4
11	402.5	13.6	1416	1	US-07-621-091G-1
12	402.5	13.6	1416	2	US-08-399-889-1

13	402.5	13.6	1416	3	US-09-167-364-1	Sequence 1, Appli
14	402.5	13.6	1416	4	US-09-439-897-1	Sequence 1, Appli
15	400.5	13.5	4359	4	US-09-484-970B-4	Sequence 4, Appli
16	399	13.4	1881	4	US-09-029-348-20	Sequence 20, Appli
17	394.5	13.3	1560	2	US-08-794-795-5	Sequence 5, Appli
18	394.5	13.3	1560	4	US-09-249-200-5	Sequence 5, Appli
19	394	13.3	5102	1	US-08-494-168-1	Sequence 1, Appli
20	393.5	13.3	503	4	US-09-297-269-40	Sequence 40, Appli
21	391.5	13.2	1560	4	US-09-453-702B-264	Sequence 264, App
22	391.5	13.2	61663	4	US-09-453-702B-62	Sequence 62, Appl
23	389.5	13.1	1703	2	US-08-794-795-1	Sequence 1, Appli
24	389.5	13.1	1703	4	US-09-249-200-1	Sequence 1, Appli
25	380.5	12.8	1868	1	US-08-392-367B-1	Sequence 1, Appli
26	380.5	12.8	1868	3	US-08-893-467A-1	Sequence 1, Appli
27	356	12.0	1839	1	US-08-383-744-1	Sequence 1, Appli
28	356	12.0	1839	2	US-08-999-336-1	Sequence 1, Appli
29	356	12.0	1839	5	PCT-US96-01427-1	Sequence 1, Appli
30	355.5	12.0	2383	4	US-09-523-487-9	Sequence 9, Appli
31	355.5	12.0	2409	3	US-09-320-095-9	Sequence 9, Appli
32	350	11.8	1458	4	US-09-111-470-3	Sequence 3, Appli
c 33 343 11.6 506 1						Sequence 61, Appli
c 34 331.5 11.2 2824 2						Sequence 3, Appli
c 35 330.5 11.1 392 4						Sequence 372, App
c 36 329 11.1 9827 4						Sequence 66, Appli
c 37 327 11.0 48908 4						Sequence 137, App
c 38 325 10.9 45175 4						Sequence 116, App
c 39 323.5 10.9 3404 4						Sequence 94, Appli
c 40 322 10.8 1370 4						Sequence 9, Appli
c 41 315.5 10.6 2851 4						Sequence 1, Appli
c 42 315.5 10.6 2851 4						Sequence 3, Appli
c 43 313 10.5 2028 1						Sequence 1, Appli
c 44 313 10.5 2028 2						Sequence 1, Appli
c 45 313 10.5 2028 2						Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-642-255-60
; Sequence 60, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph A.
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

QY 393 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 406
Db 1424 GAACAAGGCCCTGGCTGCCCC--GGATTCCAGGGTCTCCCTGGTCTCT 1480
QY 407 -----ProGluAspAsnGlyCysPro 413
Db 1481 CCAGGTGAAGCAGCAACCTGGTGAACAGGGTGTCTCT 1519

RESULT 3
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESS: FLHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642.255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20.015
; REFERENCE/DOCKET NUMBER: A5556-3/BR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-50

Alignment Scores:
Pred. No.: 2.46e-17 Length: 756
Score: 448.50 Matches: 99
Percent Similarity: 52.70% Conservative: 18
Best Local Similarity: 44.59% Mismatches: 85
Query Match: 15.11% Indels: 21
DB: 1 Gaps: 5

US-09-763-712A-2 (1-547) x US-08-642-255-50 (1-756)
QY 197 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMet-----ArgSerArg 214
Db 56 GTGGCGCTGGACCGGTGTCTCCACCGGGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGC 115
QY 215 LeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetIysLeuValAspSer 234
Db 116 CTGGACCGCTGTCTCACCGG-----GTGCTCCGGGACCTGCAGGCCCGCCAGGFG 166
QY 235 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGly 254
Db 167 CGCCTGGACCGCTGTCTCACCGG-----GTGCTCCG--GGACCTGCAGGCCCGCCAGGT 219

QY 255 ProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLys 274
Db 220 GGCCTTGACCGCGGTGCTCCACCGGTGCTCCGGACCTGCAGGCCCGCCAGGTGCGCCT 279
QY 275 GlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGlyProLysGlyPro 294
Db 280 GGACCGGTGCTCCACCGGTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGACCG 339
QY 295 AlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGly 314
Db 340 GCTGTCTCACCGGTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGACCGGTGCT 399
QY 315 SerArgGlySerPro-----GlyLysProGlyProGlnGlyProSerGlyAspPro 331
Db 400 CCACCGGTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGACCGGTGCTCCACCG 459
QY 332 GlyProGlyProGlyGlyGluGlyLeuProGlyProGlyProGlnGlyProGlyPhe 351
Db 460 GGTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGACCGGTGCTCCACCGGTGCT 519
QY 352 GlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGly 371
Db 520 CCGGACCTGCAGGCCCGCCAGGTGCGCCTGACCGGTGCTCCACCGGTGCTCCGGGA 579
QY 372 LeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyProSer 391
Db 580 CCGGACCGCCCGCGGTGCTGCAGCGGTGCTGCAGCGGTGCTCCACCGGTGCTCCGG 639
QY 392 GlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 411
Db 640 GGC-----CGCCAGGTGCGCCTGACCGGTGCTCCGGGTGCT 669
QY 412 CysPro 413
Db 670 CCACCG 675

RESULT 4
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029, 348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-19

Alignment Scores:
Pred. No.: 2.54e-16 Length: 1608
Score: 436.50 Matches: 102
Percent Similarity: 47.06% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 67
Query Match: 14.70% Indels: 70
DB: 4 Gaps: 9

US-09-763-712A-2 (1-547) x US-09-029-348-19 (1-1608)
QY 235 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGly 254
Db 59 CAACATGCCCAATCTTTACAGAGGAA-CTGTAAAGAAAGGCCCGCCAGGATAGAGA 117
QY 255 ProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLys 274

```
Db 118 CCAGTGGAGAAAGGGTCCACAGGCCCCCGCAGGACAGATGGTGAAGATGGTCCACCA 177
Qy 275 GlyGluLysGluProGlyProGlyPro----- 285
Db 178 GGCCTCTGGTCCACCTGCTCTCTGCGCCCGCTCGGTGGGAACCTTTGCTGCT 237
Qy 286 -----AlaGlyGluArgGlyProLeGlyProAla 295
Db 238 CAGTATGATGAAAGAGAGTTGGACTTGGCCCTGGACCAATGGCTTAATGGACCTAGA 297
Qy 296 GlyProGlyGluArgGlyGlyLysSerLysSerLysSerLysProLysGlySer 315
Db 298 GGCCTCTGGTCCACCTGAGGCCCGCAGGCTCAAGGTTTCAAGGACCTGCTGTGAG 357
Qy 316 ArgGlySerProGlyLys-----ProGlyProGlnGlyPro----- 327
Db 358 CTGGTGAACCTGGTCAACCTGGTCTGCAGGTGCACCTGGTCTCATGGCCCGCTGGGT 417
Qy 328 -----SerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
Db 418 CTGCTGGCAACATGAAACCGTGTGAACCTGCTCTCTGCTGCTGCTGCTGCT 477
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 478 GGTGCTGTGGCCCAAGAGGTCTAGTGGCCCAAGGACCTGCTGGCGCATAAAGGAGAG 537
Qy 361 ProGlyValProGlyProArgGlyLeuPro-----Gly 371
Db 538 CCGGGTGAAGAGGGGCCAGAGGTCTCTCTGGCTTCAAGGGACACAATGGATGGCAAGGT 597
Qy 372 LeuProGlyVal-----ProGlyMetProGlyProLys 382
Db 598 CTGCTGGTATCGTCTGCTCACCATGCTGATCAAGGTGCTCTGCTGCTGCTGCTGCT 657
Qy 383 GlyProGlyProGlyProGlyProSerGly----- 392
Db 658 GGTCTAGGGGGCCCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
Qy 393 ---AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 411
Db 718 CTGGTACGGTGGACCTGCTGG-CATTCGAGGCCCTCAGGTCACCAAGGCC----- 770
Qy 412 Cys---ProPro-----HisTrpLysAsnPhetheAspLysCys 423
Db 771 TGTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815

RESULT 5
US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
```

```
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-159-784-1

Alignment Scores:
Pred. No.: 1,99e-15 Length: 4031
Score: 428.00 Matches: 95
Percent Similarity: 46.41% Conservative: 15
Best Local Similarity: 40.08% Mismatches: 49
Query Match: 14.42% Indels: 78
DB: 1 Gaps: 8

US-09-763-712A-2 (1-547) x US-08-159-784-1 (1-4031)
Qy 231 LeuValAspSerLysHisGlyGlnLeuLeuLysAsnPhethrIleLeuGlnGlyProPro 250
Db 976 GTGGTCCAGAGCCCCAACACTCACAACTGCTCCCTGGA-----GCACAAGGACCCCG 1026
Qy 251 GlyProArgGlyPro----- 255
Db 1027 GGACCTCAGGGGCCACCAGGAGGATGGCACTCCAGGAAGGATGGTGAACCGGGTGAC 1086
Qy 256 -----ArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGly 269
Db 1087 CCTGGTGAAGATGGGAGACCGGGTGACACTGGACCTCAAGGCTTTCAGGAGCCCGGCA 1146
Qy 270 AsnLysGlyGlnLysGlyGluLysGlyGluPro-----GlyProProGly 284
Db 1147 GATGTGGCCCTCAAGGCGAGAGGAGATCTCTGTTGGGCCCGCCGAGGACCTCCAGGG 1206
Qy 285 ProAlaGlyGluArgGlyProLe----- 292
Db 1207 CCTCCAGGGCCACCAGGACCTCTCTTCAGACAAGACAAGCTGACCTTCATTGACATGGAG 1266
Qy 293 -----GlyPro 294
Db 1267 GGATCCGGTTTCAGCGGAGACATAGAGAGCCTTAGAGGCCACAGAGGCTTCCCTGGCCCC 1326
Qy 295 AlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGly 308
Db 1327 CCGGGGCCCGCTGGTGTCCAGGACTTCTGTGACCCAGGACGCTTGGGATCAATGGT 1386
Qy 309 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 328
Db 1387 TCCTATGCACCA---GGACCTGCAGGCCCTTCCTGTGTACCTGGGAAGGAGGACCCCG 1443
Qy 329 GlyAspProGlyProProGlyProLysGluGlyLeuProGlyProGlnGlyPro 348
Db 1444 GGTTCAGGTCCTCCCGGACCTCCAGGTCCT-----CCAGGCAAGAGGCGCCCA 1494
Qy 349 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProAsnG--- 367
Db 1495 CCAGGAGTGGCGGCCAGCAAGAGGACGTGTGTGTGTGGGATGCCAGGACCAAGGG 1554
Qy 368 -----GlyLeuProGlyLeuProGlyValProGlyMetPro 379
Db 1555 AGCAAGAGGACCTTGGGGCCCATCGTATGCTGTGGCAAGCTGCTGGCTGGTGGTCCCT 1614
Qy 380 GlyProLysGlyProProGlyProGlyProGlyProGlyProGlyProGlyPro 396
```


Db 1615 GGGCCAGTTGGACCCCGAGGACCTCCAGGGCTCCAGGGCCACGAGGACCA 1665

RESULT 6

US-09-297-269-39

; Sequence 39, Application US/09297269

; Patent No. 6451557

; GENERAL INFORMATION:

; APPLICANT: VAUGHAN, Paul R.

; APPLICANT: GALANIS, Maria

; APPLICANT: RAMSHAW, John A.M.

; APPLICANT: WERKMEISTER, Jerome A.

; TITLE OF INVENTION: A METHOD FOR PRODUCING, IN YEAST, A HYDROXYLATED TRIPLE

; TITLE OF INVENTION: HELICAL PROTEIN, AND YEAST HOST CELLS USEFUL IN SAID

; TITLE OF INVENTION: METHOD (As Amended)

; FILE REFERENCE: Q54094

; CURRENT APPLICATION NUMBER: US/09/297,269

; CURRENT FILING DATE: 1999-04-28

; EARLIER APPLICATION NUMBER: P03310

; EARLIER FILING DATE: 1996-10-29

; EARLIER APPLICATION NUMBER: P04306

; EARLIER FILING DATE: 1996-12-19

; EARLIER APPLICATION NUMBER: PCT/AU97/00721

; EARLIER FILING DATE: 1997-10-29

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 39

; LENGTH: 1572

; TYPE: DNA

; ORGANISM: Collagen Type III Alpha I Chain

US-09-297-269-39

Alignment Scores:

Pred. No.:	1.14e-15	Length:	1572
Score:	424.50	Matches:	119
Percent Similarity:	42.48%	Conservative:	25
Best Local Similarity:	35.10%	Mismatches:	108
Query Match:	14.30%	Indels:	88
DB:	4	Gaps:	15

US-09-763-712A-2 (1-547) x US-09-297-269-39 (1-1572)

QY 224 ValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhe 243

Db 194 GTGAACCTGGAGAGATGGAACCCCTGGAT-----CAGATGGTC-----TTC 235

QY 244 ThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer----- 261

Db 236 CAGGTGCA-GATGGATCTCTGGTGCAGAGGTGATCGTGGTGAATGGCTCTCTCTGGT 294

QY 262 -----GlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLys 274

Db 295 GCCCCTGGGGCTCTGGTCAATCCAGGCCACCTGGTCTCTGGTCCAGCTGGAAGAGT 354

QY 275 GlyGluLysGlyGluProGlyPro-----ProGlyProAlaGlyGlu 288

Db 355 GTGACAGAGAGAGAAGTGGCCCTGCTGCCCTGTGTGCTGCTGCTGCTGCTGCTGCT 414

QY 289 ArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGly 308

Db 415 CGAGGTGCTCTGGTCTCAAGGCCACCTGGTGAAGAGGTGAACAGGTGAACGTGGA 474

QY 309 SerGlnGlyProLysGlySerArg-----GlySerProGlyLysPro 322

Db 475 GCTGCTGGCATCAAGGACATCGAGGATTCCTGTAATCCAGGTGCCCGAGTTCTCCA 534

QY 323 GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeu 342

Db 535 GGCCCTGTGTGTCAGCAGGGTGAATCGGCAGTCCAGGACCTGCAGGCCCGCCAGAGACCT 594

QY 343 ProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly 362

Db 595 GTTGACCCAGTGCCTCTCGCAAGAGATGGAACAGTGGACATCCAGGTCCCATTTGGA 654

QY 363 ValProGlyPro-----ArgGlyLeuProGlyLeuProGlyValPro 376

Db 655 CCACCAGGGCCTCGAGGTAACAGAGGTGAAGAGGATCTCAGGGCTCCCGCCAGGCCACCA 714

QY 377 GlyMetProGlyProGlyGlyProGlyProGlyProGlyProGlyProSer-----Gly 392

Db 715 GGGCAACAGGCCCTCTGGACCTCTCTGGTGGCCCTGGTGGTGGTGGTGGTGGTGGT 774

QY 393 AlaValValProLeuAlaLeuGlnAsnGlu-----ProThrProAla 406

Db 775 GCGCTGCCATTCCTGGGATTCGAGGTGAAGAGCTGGCGGTTTCCCGCTATTATGGA 834

QY 407 ProGluAspAsnGlyCysProHisTrpLysAsnPheThrAsp----- 421

Db 835 CCTGAA-----CCAATGGATTTCAAAATCAACACCGATGAGATATCACTTCA 882

QY 422 ---LysCysTyTyTyPheSerValGluLysGluIlePheGluAspAlaLys----- 437

Db 883 CTCAAGTCTGTTAATGGACAAATAGAAAGCCCAATAGTCTCTGATGGTCTCTCGTAAAC 942

QY 438 -----LeuPheCysGluAspLys----- 443

Db 943 CCGCTAGAACTGCAGAGACCTGAATTCCTGCCATCTCAAGCTCAAGACTGCAGATAC 1002

QY 444 -----SerSerHisLeuValPheIleAsnThrArg 453

Db 1003 TGGGTGCACCTTAACCAAGGATGCAAAATTTGGATGCTATCAAGGTATCTGTAATATGGA 1062

QY 454 GluGluGlnGlnTrpIleLysLysGlnMetValGly---ArgGluSerHisTrpIleGly 472

Db 1063 ACTGGGAAACATGCAATAGTCCCAATCTTTGAATGTTCCACGGAACACTGG----- 1116

QY 473 LeuThrAspSerGluArgGluAsnGluTrpLysTrp-----LeuAspGly 487

Db 1117 TGCACAGATTCAGTCTGCTGAGAGAAACACAGCTTTGGTTGGAGAGTCCATGATGCT 1173

RESULT 7

US-08-555-669-11

; Sequence 11, Application US/08555669

; Patent No. 5773248

; GENERAL INFORMATION:

; APPLICANT: Brewton, Richard G.

; APPLICANT: Mayne, Richard

; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/555,669

; FILING DATE: 13-NOV-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Halluin, Albert P.

; REGISTRATION NUMBER: 25,227

; REFERENCE/DOCKET NUMBER: 8389-030

; TELEPHONE: 415-854-3660

; TELEFAX: 415-854-3694

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2543 base pairs

; TYPE: nucleic acid

```
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-08-555-669-11

Alignment Scores:
Pred. No.: 3,85e-15      Length: 2543
Score: 419.00           Matches: 121
Percent Similarity: 35.59%      Conservative: 26
Best Local Similarity: 29.30%    Mismatches: 91
Query Match: 14.11%           Indels: 176
DB: 1                      Gaps: 9

US-09-763-712a-2 (1-547) x US-08-555-669-11 (1-2543)

QY 232 ValAspSerLysHisGlyGlnLeuIleLysAsnPhetHrIleLeuGlnGlyProGly 251
DB 986 CTCGATGGCCAGGAAGGAGAGGCTGGTCCGAAC-----GGTGCTCGGGA 1030
QY 252 ProArgLysGlyArgGlyAspArgGlySerGlnGlyProGlyProGlyProGly 271
DB 1031 GAGAAGGGCCCCAACGGGTGCCCGGCTCCCTGGAGAGCGGGGTCCAAAGCGGAGAAG 1090
QY 272 GlyGlnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyPro 291
DB 1091 GGAGACGGGGCAGAGCTGGGAGCTGGGTGAGGCGCGCCCTCTGGAGAGCCAGCGCTC 1150
QY 292 IleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer----- 309
DB 1151 CTTGGAGATGCTGGCATGCTGGGGAGCGCGGTGAGGCTGCCACCGGGCTCAGCGGG 1210
QY 310 -----GlnGlyProLys 313
DB 1211 GCCTCGGCCCAAGCCCTCCCGAGGCCCTGGTGTCCGAGGCTTCAGGGCCCAAG 1270
QY 314 GlySerArgGlySerProGlyLysProGlyProGlnGlyProGlyProGlyProGlyPro 333
DB 1271 GGCAGCATGGAGACCCCGGCTTCAGGCCCCCAGGCTCCGAGGTGACGTGGCGCAG 1330
QY 334 ProGlyProGlyLysGlnGlyLeuProGlyProGlyProGlnGlyProGlyPheGlnGly 353
DB 1331 CGGGTCCGGGAGTCCCAAGGCCCTTAAGGAGGACACAGGATTATTCAGGTTCCGACGT 1390
QY 354 LeuGlnGlyThrValGlyGlu-----GlnGlyProLys 360
DB 1391 CTTCTGGGATAAAGAGAACTGGGTCCCGAGGCTCGTCGGACCCCAAGAGAGTCT 1450
QY 360 ----- 360
DB 1451 GGCAGTCGAGGGAGCTGGGCCCCCAAGGACCCAGGCTCCCAAGCGCACCGAGCTGT 1510
QY 361 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 377
DB 1511 CAGGGTGTCCCGGGCCCCCGGCTCCTGGGCTTCAGAGGCTCCCGGGTGTCTCTGTC 1570
QY 378 Met----- 378
DB 1571 ATCAGCGGAAGCCGGAGTTCGGGGAAGGAGGCCAGCAGCAGCGCATCAGGAGCTG 1630
QY 378 ----- 378
DB 1631 TGTGGGGGATGATCAGCGAACAATAATTCACAGTTAGCGCGCACCTAAGGAAGCCTTTG 1690
QY 379 -----ProGlyProLysGlyProGlyProGlyProGlyProGlyProSer 391
DB 1691 GCACCGGGTCCATTGGTCCGGCCCGGTCCAGTGGCCCTCGGCGCCCGCCAGGACCCCA 1750
QY 392 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 403
DB 1751 GGCTCATTTGGTCACCTCGCGCTCGAGGACCCCGCGGATACCGCGGTCCCTCCAGTGGGAG 1810
```

```
QY 404 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPhetHras 421
DB 1811 CTGGGAGACCCCGGCCAGAG----- 1832
QY 421 pLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG1 441
DB 1833 -----GAAACCCAGGTTGACAGAGAGACAAG----- 1859
QY 441 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLy 461
DB 1860 -----GCGCGCAGAGCAG----- 1874
QY 461 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG1 481
DB 1875 -----GGCTGGAGCGGCTTGAAGGAGACACGAGGCCCCCAAGGA 1912
QY 481 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAs 501
DB 1913 CCCCAGAGC-GTCCCGGACCAAGCAAGGACGCGGCGGAGCGTCTCCCGCGAGCTGG 1971
QY 501 pAsnTrpGlyHisGlyHisGlyProGly----- 510
DB 1972 GCCTCCCGGAGATCTGGGCTTCCAGGTGCCATGTCGAGGCGGACACCGCGGATCTG 2031
QY 511 -----GluAspCysAlaGlyLeuIleTyrAlaGly 520
DB 2032 CGACACCTCAGCTGCCAAGGACCGCTGTAGGAGG 2068

RESULT 8
US-09-073-663-11
; Sequence 11, Application US/09073663
; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,663
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-09-073-663-11
```

Alignment Scores:			
Pred. No.:	3.85e-15	Length:	2543
Score:	419.00	Matches:	121
Percent Similarity:	35.59%	Conservative:	26
Best Local Similarity:	29.30%	Mismatches:	91
Query Match:	14.11%	Indels:	176
DB:	3	Gaps:	9
US-09-763-712A-2 (1-547) x US-09-073-663-11 (1-2543)			
QY	232	ValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGly	251
DB	986	CTCATATGGCCAGAGGAGGCTGTCTCCACAC-----GGTCTCCGGGA	1030
QY	252	ProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys	271
DB	1031	GAGAAAGGGCCCAACGCGGTGCGGGCCCTCCCTGGACGAGCGGGTCCAAAGCGGAGAG	1090
QY	272	GlyGlnLysGlyGlnLysGlyGlnProGlyProProGlyProAlaGlyGlnArgGlyPro	291
DB	1091	GGAGAACGGGGCAGAGCTGGGAGCTGGGTGAGCGCGGCCCTCTGGAGAGCCAGCGCTC	1150
QY	292	IleGlyProAlaGlyProProGlyGlnArgGlyGlyLysGlySerLysGlySer-----	309
DB	1151	CCTGGAGATGCTGGCATGCTTGGGGAGCGCGGTGAGCTGGCCACCGGGCTCAGCGGGG	1210
QY	310	-----GlnGlyProLys	313
DB	1211	GCCTCGGGCCCAAGCCCTCCCGAGCCCTGGTGTCCGAGGCTTCAGGGCCAGAG	1270
QY	314	GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro	333
DB	1271	GGCAGCATGGGAGACCCCGCGCTTCAGAGCCCGCCAGGGCTCCGAGGTGACGTGGCGGAC	1330
QY	334	ProGlyProProGlyLysGlnGlyLeuProGlyProGlnGlyProProGlyPheGlnGly	353
DB	1331	CGGGGTCCGGAGGTCCTTAAGGAGCCCTAAGGAGACCGAGGTATTCAGGTTCCCGACGT	1390
QY	354	LeuGlnGlyThrValGlyGlu-----	360
DB	1391	CTTCTGGGATTAAGAGACACTGGTCCAGCGGCTGTTCGGACCCCAAGGAGAGTCT	1450
QY	360	-----	360
DB	1451	GGCAGTCGAGGGAGCTGGGCCCCCAAGGACCCAGGGTCCCAAGCGCACCGCGTGT	1510
QY	361	-----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly	377
DB	1511	CAGGGTGTCCCGGGCCCCCGCTCTCTGGGCTTCAGGGGCTCCCGGGTGTCTCTGGC	1570
QY	378	Met-----	378
DB	1571	ATCACGGGAGCCGGAGTTCGGGGAGGAGGAGCCAGCAGCGCATCAGGGAGCTG	1630
QY	378	-----	378
DB	1631	TGTGGGGGATGATCAGCAACAAATTGCACAGTTAGCCGCGCACCTAAGGAAGCCTTTG	1690
QY	379	-----ProGlyProLysGlyProProGlyProProGlyProGlyProser	391
DB	1691	GCACCGGGTCCATTGGTGGCGCCCGGTCCAGTGGCCCGCCCGCCCGCCAGGACCCCA	1750
QY	392	GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro-----	403
DB	1751	GGCTCATTTGGTCACCTCGGCTCGAGGACCCCGCCGATACCGCGGTCCCTACTGGGGAG	1810
QY	404	-----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs	421
DB	1811	CTGGGAGACCCCGGGCCAGAG-----	1832
QY	421	pLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG	441
DB	1833	-----GAAACACGAGGTGACAGAGAGACAAAG-----	1859

QY 441 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLy 461
 DB 1860 -----GGCGCGGAGGAGCAG----- 1874
 QY 461 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG 481
 DB 1875 -----GGCTGGACGGGCTCAAGGAGCAGGAGGCGGCGCCCAAGGA 1912
 QY 481 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAs 501
 DB 1913 CCCCAGGC-GrCCCCGCGCACCAAGGAGCGCCAGGAGGGTGTCTCCCGCGGAGCTGTG 1971
 QY 501 pAsnTrpGlyHisGlyHisGlyProGly----- 510
 DB 1972 GCCTCCCGGAGATCTCTGGCTTCCAGGTGTCATTGGGCGCCAGGAGCGGATCTG 2031
 QY 511 -----GluAspCysAlaGlyLeuIleLysAlaGly 520
 DB 2032 CGACACTCAGCTCGCTGCAAGGAGCGCTGTAGGAGGG 2068
 RESULT 9
 US-08-642-255-48
 ; Sequence 48, Application US/08642255
 ; Patent No. 5773249
 ; GENERAL INFORMATION:
 ; APPLICANT: CAPPELLO, Joseph
 ; APPLICANT: FERRARI, Franco A.
 ; TITLE OF INVENTION: High Molecular Weight Collagen-Like
 ; TITLE OF INVENTION: Protein Polymers
 ; NUMBER OF SEQUENCES: 135
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/642,255
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROWLAND, Bertram I.
 ; REGISTRATION NUMBER: 20,015
 ; REFERENCE/DOCKET NUMBER: A55556-3/BIR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 494-8700
 ; TELEFAX: (415) 494-8771
 ; TELEX: 910 277299 FHT UR
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 432 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: other nucleic acid
 ; DESCRIPTION: /desc = "synthetic"
 US-08-642-255-48

Alignment Scores:
 Pred. No.: 1.34e-15 Length: 432
 Score: 412.50 Matches: 80
 Percent Similarity: 55.70% Conservative: 3
 Best Local Similarity: 53.69% Mismatches: 57
 Query Match: 13.89% Indels: 9
 DB: 1 Gaps: 3

US-09-763-712A-2 (1-547) x US-08-642-255-48 (1-432)

```

QY 248 GlyProGlyProArgGlyProArgGlySerGlnGlyProProGlyPro 267
Db 10 GGGCCAGCAGTCCGAGGGCGCGATGTCAGCAGGCCCCGAAAGGTCCGCTGACCG 69
QY 268 ThrGlyAsnLysGlnLysGluLysGluProGlyProProGlyProProLagly 287
Db 70 GCTGGTCCACCGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGT 129
QY 288 GluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlySerLys 307
Db 130 CCACCGGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGTCCACCG 189
QY 308 GlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 327
Db 190 GGTGCTCCGGGACCT-----GCAGGCCCGCCAGGTGCGCTGGACCGGTGTCCA 240
QY 328 SerGlyAsnProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGly 347
Db 241 CCGGGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGT 291
QY 348 ProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 367
Db 292 CCACCG-----GGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGT 342
QY 368 GlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyPro 387
Db 343 GGTCCACCGGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGT 402
QY 388 ProGlyProSerGlyAlaValValPro 396
Db 403 AAGGAGCTCAGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGT 429

```

RESULT 10

US-08-159-784-4

```

; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4
Alignment Scores:
Pred. No.: 1.87e-14 Length: 3394
Score: 409.00 Matches: 86
Percent Similarity: 48.98% Conservative: 10
Best Local Similarity: 43.88% Mismatches: 42
Query Match: 13.78% Indels: 58
DB: 1 Gaps: 7
US-09-763-712A-2 (1-547) x US-08-159-784-4 (1-3394)
QY 246 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySer----- 261
Db 553 ATGCCCGGGCCCCCAGGACCTCCAGGGCCCCCAGGCCCCCTCCAGGGACTCTGTTTACGAC 612
QY 262 -----GlnGlyProProGlyProProGlyProProGlyProProGlyProProGlyPro 272
Db 613 AGCAATGTGTTGCTGAGTCCAGCGCCCGCGGCTCCAGGATTGCCAGGAATCAG--- 669
QY 273 GlnLysGlyGluLysGlyGluProGlyProProGlyProProGlyProProGlyProProGly 292
Db 670 -----GGCCCTCCAGGACCCCAAGGGCCCCCAAGAGGAAGTG 705
QY 293 GlyProAlaGlyProProGlyGlu----- 300
Db 706 GGGCCCCCGGACCCAGGCGAGTTTCCGTTTGACTTCTTCAGAGGAGGCTGAATG 765
QY 301 ArgGlyGlyLysGlySerLysGlnGlyProLysGlySerArgGlySer----- 318
Db 766 AAGGGGAGAGGAGACCGAGGTGATGACGAGACAGAAAGGGGGAGGCGCGG 825
QY 319 -----ProGlyLysProGlyProGlnGlyProSerGly 329
Db 826 GGGCGGGTTCCTTCGGCTCCAGCGTCCGGGGCCCCCGGGCCCCCAGGCCCCCAGCTGG 885
QY 330 AspProGlyProProGlyProGlyLysGlu-----GlyLeuProGly 344
Db 886 TACCCTGGGATTCAGTCCCAAGGGAGAGCATCCGGGGCCAGCGCCGCCACCTGGA 945
QY 345 ProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 364
Db 946 CCTCAGGGACCCCGCGCATC---GGCTACGAGGGGCGCAGGGCCCTCCCGGCCCCCA 1002
QY 365 GlyProArgGlyLeuProGlyLeuProGly-----ValProGly 377
Db 1003 GGGCCCCCAGGGCCCCCTTCATTTCTTCGCGCTCCAGGACGACTATCAGCGTTCCGGG 1062
QY 378 MetProGlyProLysGlyProProGlyProProGlyProSerGlyAla 393
Db 1063 CCTCGGGGGCCCCCTGGGGCCCCCTGGGGCCCCCTGGAACCATGGCGCC 1110

```

RESULT 11

```

US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage

```

; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh OS7.0
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/621,091G
 ; FILING DATE: 11/30/90
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA: No. 5424408 applicable
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Barth, Richard S.
 ; REGISTRATION NUMBER: 28180
 ; REFERENCE/DOCKET NUMBER: 900983/RB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 972-1400
 ; TELEFAX: (212) 370-1622
 ; TELEX: 236268
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1416 base pairs
 ; TYPE: Nucleic acid
 ; STRANDEDNESS: Double
 ; TOPOLOGY: Linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHETICAL: No
 ; ANTI-SENSE: No
 ; ORIGINAL SOURCE:
 ; ORGANISM: Calf
 ; STRAIN: Unknown
 ; INDIVIDUAL ISOLATE: Unknown
 ; DEVELOPMENTAL STAGE: Unknown
 ; CELL TYPE: Whole kidney
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Bovine lens cDNA
 ; CLONE: KMC15
 ; POSITION IN GENOME: No. 5424408 known
 ; CHROMOSOME/SEGMENT: No. 5424408 known
 ; PUBLICATION INFORMATION: No. 5424408e
 ; US-07-621-091G-1

Alignment Scores:
 Pred. No.: 1416
 Score: 1.69e-14
 Percent Similarity: 402.50
 Best Local Similarity: 46.62%
 Query Match: 13.56%
 DB: 1
 Gaps: 8

US-09-763-712A-2 (1-547) x US-07-621-091G-1 (1-1416)

QY 247 GlnGlyProGlyProArgGlyProArgGlySerGlnGlyProGly 266
 Db 112 AAAGAGATCGAGGTCCACCTGCTCCAGAGAACCCAGGCATGCTGCTCTCTGA 171
 QY 267 ProThrGlyAsn-----LysGlyGlnLysGlyGluProGly 281
 Db 172 CCTCCAGGAGGCTCTGTAGAGGCATATAAAGGAGACAGGGGTGATGGAGGCTGCG 231
 QY 282 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 301
 Db 232 CAAGAGGTCCACCTGAGCTATAGAGACATGGGGTCCAGAGTCATCCGGAGCACCA 291
 QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
 Db 292 GGTGTCCCGGTGAGGAGGAGGAGGAGGATGATGATGATGATGATGATGATGATGATG 351
 QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 341
 Db 352 AAAGGAGAGAGGGTAATTTCAGGATTTCCAGGACACCTGGACCTCCAGGCAAGTGA 411
 QY 342 -----LeuPro 343
 Db 412 CCAAAAGGACACCTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATCTCCCTTCA 471

QY 344 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
 Db 472 GGAAGCCAGGCGCCACCTGCTCAGCTGGAGAACCCAGGATGCAAGGAGAACCCGGGCC 531
 QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
 Db 532 CCAGGACACCGAGGATCCAGGACCTGTGGGCCAAAAGGTAAACCCAGGGAGGATGGT 591
 QY 384 ProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
 Db 592 CCACAGGAACTCTCTGACCACTGGAGAAAAGGCAACAAAGTTGCTAAAGGAGACAA 651
 QY 404 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
 Db 652 GGACCA---CCTGGATCCGATGCTGCCAGGCTTGAAGGGGAAACCT-TGAGACACATGG 707
 QY 420 ThrAspLysCysTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439
 Db 708 ACC---ACCTGCAGCAGGGGCGATGATGAGGGGCTTTGCTTTACCCGGGCACAG----- 758
 QY 440 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
 Db 759 ---CCAGACACCACTTCCCTCTGTCAGAGGAGGACAGACCCGCTCTATAGTGGTT 815
 QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
 Db 816 TTCTCTCTCTTTGTACAAAGAAATGAACAAGCCCATGGACAGGA----- 860
 QY 478 ArgGluAsnGluTrpLys 483
 Db 861 CCTGGGACACTTGGCAG 878
 RESULT 12
 ; US-08-399-889-1
 ; Sequence 1, Application US/08399889B
 ; Patent No. 5973120
 ; GENERAL INFORMATION:
 ; APPLICANT: Readers, Stephen T
 ; APPLICANT: Morrison, Karen E
 ; APPLICANT: Hudson, Billy G
 ; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
 ; FILE REFERENCE: 951263A
 ; CURRENT APPLICATION NUMBER: US/08/399,889B
 ; CURRENT FILING DATE: 1995-03-07
 ; EARLIER APPLICATION NUMBER: 07/621091
 ; EARLIER FILING DATE: 1990-11-30
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1416
 ; TYPE: DNA
 ; ORGANISM: Calf
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1416)
 ; US-08-399-889-1
 Alignment Scores:
 Pred. No.: 1.69e-14
 Score: 402.50
 Percent Similarity: 46.62%
 Best Local Similarity: 38.72%
 Query Match: 13.56%
 DB: 2
 Gaps: 8
 US-09-763-712A-2 (1-547) x US-08-399-889-1 (1-1416)
 QY 247 GlnGlyProGlyProArgGlyProArgGlySerGlnGlyProGly 266
 Db 112 AAAGAGATCGAGGTCCACCTGCTCCAGAGAACCCAGGCATGCTGCTCTCTGA 171
 QY 267 ProThrGlyAsn-----LysGlyGlnLysGlyGluProGly 281
 Db 172 CCTCCAGGAGGCTCTGTAGAGGCATATAAAGGAGACAGGGGTGATGGAGGCTGCG 231

Db 172 CCTCCAGGAGTCCTCTAGAGGCATATAAAGGAGACAAGGGGTGTATGGAGAGCCTGGC 231
QY 282 ProProGlyProAlaGlyLeuArgGlyProIleGlyProAlaGlyProGlyGluArg 301
Db 232 CAAAGAGTCCACCTGGAGCTATAGGAGCATGGGTACCAGGTCTATCCGGGAGACCA 291
QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
Db 292 GGTGTCCCGGTGAGCCAGCGGCGAGAGGTATCTCTGGATTCTATGGATTCCAGGCATG 351
QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGly 341
Db 352 AAAGGAAGAAGGGTAATTCAGATTTCAGGACCACTGGACCTCCAGGGCAAGGTGA 411
QY 342 -----LeuPro 343
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 344 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
Db 472 GGAAGCCCAAGGCCACCTGGTTCAGCTGGAGAACCAAGGATGCAAGGAGAACCCGGGCC 531
QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
Db 532 CCAGGACCAAGGAGATCCAGGACCTGTGGCCCAAGGTAACCAAGGAGGATGGT 591
QY 384 ProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
Db 592 CCACCAAGGAACCTCTGGACCACTGGAGAAAAGGCAACAAGGTTGTAAGGAGAGCAA 651
QY 404 ThrProAlaProGluAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
Db 652 GGACCA---CCTGGATCCGATGGCTGCCAGGCTTGAAGGGGAAACCC-TGGAGACACTGG 707
QY 420 ThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439
Db 708 ACC---ACCTGCAGCAGGGCGAGTGTATGGGGCTTGTCTTACCCTGGCAGAG----- 758
QY 440 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
Db 759 ---CCAGACCACAGCAATTCCTCTCCAGAGGAGGAGGAGCGCTCTATAGTGGTT 815
QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
Db 816 TTCTCTCTCTTTGTACAAGAAATGAACAAGCCCATGGACAGGA----- 860
QY 478 ArgGluAsnGluTrpLys 483
Db 861 CCTGGGAACACTGGCAG 878

RESULT 13
US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reenders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951262B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)

US-09-167-364-1

Alignment Scores:

Pred. No.: 1-69e-14 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 13.56% Indels: 40
DB: 3 Gaps: 8

US-09-763-712A-2 (1-547) x US-09-167-364-1 (1-1416)

QY 247 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 266
Db 112 AAAGAGATCGAGGTCCACCTGGCTCCAGAGGAACCCAGGCATGCTGGTCTCTCTCTGGA 171
QY 267 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 281
Db 172 CCTCCAGGAGTCCCTCTAGAGGCATATAAAGGAGACAAGGGGTGTATGGGAGAGCCTGGC 231
QY 282 ProProGlyProAlaGlyLeuArgGlyProIleGlyProAlaGlyProProGlyGluArg 301
Db 232 CAAAGAGTCCACCTGGAGCTATAGGAGATGGGTCCACAGGTCTATCCGGGAGCACCA 291
QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
Db 292 GGTGTCCCGGTGAGCCAGCGGCGCAGAGGTGATCCTGGATTCTATGGATTTCAGGCATG 351
QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGly 341
Db 352 AAAGGAAGAAGGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGGGCAAGGTGA 411
QY 342 -----LeuPro 343
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATCTCCCTTCCA 471
QY 344 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
Db 472 GGAAGCCCAAGGCCACCTGGTTCAGCTGGAGAACCAAGGATGCAAGGAGAACCCGGGCC 531
QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
Db 532 CCAGGACCAAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACCAAGGAGGATGGT 591
QY 384 ProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
Db 592 CCACCAAGGAACCTCTGGACCACTGGAGAAAAGGCAACAAGGTTGTAAGGAGAGCAA 651
QY 404 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
Db 652 GGACCA---CCTGGATCCGATGGCTGCCAGGCTTGAAGGGGAAACCC-TGGAGACACTGG 707
QY 420 ThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439
Db 708 ACC---ACCTGCAGCAGGGCGAGTGTATGGGGCTTGTCTTACCCTGGCAGAG----- 758
QY 440 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
Db 759 ---CCAGACCACAGCAATTCCTCTCCAGAGGAGGAGGAGCGCTCTATAGTGGTT 815
QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
Db 816 TTCTCTCTCTTTGTACAAGAAATGAACAAGCCCATGGACAGGA----- 860
QY 478 ArgGluAsnGluTrpLys 483
Db 861 CCTGGGAACACTGGCAG 878

RESULT 14

US-09-439-897-1
; Sequence 1, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:

APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1

LENGTH: 1416

TYPE: DNA

ORGANISM: Bos taurus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1416)

US-09-439-897-1

Alignment Scores:
Pred. No.: 1416
Score: 1.69e-14
Length: 1416
Matches: 103
Percent Similarity: 46.62%
Conservative: 21
Best Local Similarity: 38.72%
Mismatch: 103
Query Match: 40
DB: 8
Gaps: 4

US-09-763-712a-2 (1-547) x US-09-439-897-1 (1-1416)

QY 247 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 266
DB 112 AAGGAGATCGAGTCCACTGCTCCAGAGGAACCCAGGATGCTGCTCTCTGGA 171
QY 267 ProThrGlyAsn-----LysGlnGlnLysGlyGluLysGlyGluProGly 281
DB 172 CCTCCAGGAGTCTGTAGAACGATATAAAGGAGACAAAGGGTTGATGGAGAGCCTGTC 231
QY 282 ProGlyProAlaGlyGluArgGlyProGlyProGlyProAlaGlyProGlyGluArg 301
DB 232 CAAAGAGGTCACCTGGAGCTATAGAGACATCCAGGATCATCCGGAGACCA 291
QY 302 GlyGlyLysSerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
DB 292 GGTGTCCCGGTGAGCCAGGCGGAGAGGTGATCTGATGATTTCCAGGCATG 351
QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGly 341
DB 352 AAAGGAAGAGAGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGGGCAAAAGTGA 411
QY 342 -----LeuPro 343
DB 412 CCAAAAGGACCACTGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATCTCCCTTCCA 471
QY 344 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
DB 472 GGAAGCCAGGCCCCACCTGTTGAGTGGAGAACAGGATGCAAGGAAACCCGGGCC 531
QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
DB 532 CCAGGACCAACAGGAGATCCAGGACCTGTGGGCCCAAAAGGTAACACAGGGAGGATGT 591
QY 384 ProGlyProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 403
DB 592 CCACCAAGGACCTCTGGACCACTGGAGAAAAGGCAACAAAGTTGTAAAGAGAGAGCAA 651
QY 404 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
DB 652 GGACCA--CCTGGATCCGATGCGCTGCGAGGCTTGAAGGGGAACCT--TGGAGACACTGG 707
QY 420 ThrAspLysCysTyrTyrPheSerValGluLysGluPheGluAspAlaLysLeuPhe 439
DB 708 ACC---ACCTGACGACGAGGCGAGTGTAGAGGCGCTTTGTCTTTACCCGACACAG----- 758
QY 440 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
DB 759 ----CCAGACCAACGCAATTCCTCTCTGCTCCAGAAAGGACAGAGCCGCTCTATAGTGGT 815

QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
DB 816 TTCTCTCTCTCTTTGTACAGGAATGAACAAGCCCATGGACAGCA-----860
QY 478 ArgGluAsnGluTrpLys 483
DB 861 CCTGGGAACACTTGGCAG 878

RESULT 15

US-09-484-970B-4

Sequence 4, Application US/09484970B

Patent No. 6426186

GENERAL INFORMATION:

APPLICANT: Jones, Karen A.

APPLICANT: Volkman, Wayne

TITLE OF INVENTION: BONE REMODELING GENES

FILE REFERENCE: PB-0014 US

CURRENT APPLICATION NUMBER: US/09/484,970B

CURRENT FILING DATE: 2000-01-18

NUMBER OF SEQ ID NOS: 172

SOFTWARE: PERL Program

SEQ ID NO 4

LENGTH: 4359

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. 6426186 331365 (1553795CB1)

US-09-484-970B-4

Alignment Scores:
Pred. No.: 7,21e-14
Score: 400.50
Length: 4359
Matches: 90
Percent Similarity: 54.89%
Conservative: 11
Best Local Similarity: 48.91%
Mismatch: 64
Query Match: 13.49%
Indels: 19
DB: 4
Gaps: 7

US-09-763-712a-2 (1-547) x US-09-484-970B-4 (1-4359)

QY 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArg-----259
DB 448 CTGCCAGGACCCAGGCCCCAGGTCGCAAGAGGCTCCGAGGAATCCAGGCTTCGCA 507
QY 260 GlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGlu 279
DB 508 GGAGCTGTAGTGGAGACCAAGGCTCCAGGGGCTTCCAGGAGACCGAGGTCTGTGAAGGTT 567
QY 280 ProGlyProPro-----GlyProAlaGlyGluArgGlyProGlyProAlaGly 296
DB 568 CCAGGACCCAGGTTTCATAGGACCCCGAGGATCCAAAGTGCAGTGGGCTCCCTGGC 627
QY 297 ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArg 316
DB 628 CCAGATGGATCCCAAGTCCCATCGGCTGCCAGGGCCAGATGGGCCCTCGGGGAAGG 687
QY 317 GlySerProGlyLysPro---GlyProGln---GlyProSerGlyAspProGlyProPro 334
DB 688 GGCCTCCCTGGAGAGTCTGTGGAGTCTAGCCCGGCGCCAGGGGATGCTGTGTGTGCT 747
QY 335 GlyProGlyLysGluGlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeu 354
DB 748 GGACACCTGGGCTTAAGGCTTCCCGAGACAGAGGCCCCCTGGATTCAGAGGAGC 807
QY 355 GlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyProGly 374
DB 808 CAAGGATG-----CCTGGGATGTCAGGGCTGAAGGGCCAGCCAGGCTCCAGGA 858
QY 375 ValProGlyMetProGlyProLysGlyProGlyProProGlyPro-----ProGlyProSer 391
DB 859 CCTTCCGGCCAGCCAGGCTGTATGGGCTTCCAGGACTGCATGATGATTCAGGAGCTCCT 918

```

QY 392 GlyAlaValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaProGluAsp 409
Db 919 GGCCAAAGAGGGCCCTTGGGGCTGCCAGGAATCCACAGGCCGTGAAGGTCTGCCTGGTGAT 978
QY 410 AsnGlyCysPro 413
Db 979 AGAGGGGACCCCT 990

```

Search completed: March 21, 2003, 12:34:04
Job time : 73.1137 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:23:12 ; Search time 265.779 Seconds
(without alignments)
1596.010 Million cell updates/sec

Title: US-09-763-712A-2

Perfect score: 2969

Sequence: 1 MYSHNVVIMNLLNLTVQV.....EDVNNFTCEKDRVTLSSAL 547

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09763712/runat_14032003.100950.18119/app_query.fasta_1.1877
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09763712.cgn1.1.123 @runat_14032003.100950.18119
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_NA:
1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2936	98.9	2930	10	US-09-745-763-198
2	474.5	16.0	1410	10	US-09-954-456-782
3	455	15.3	1133	9	US-09-924-340-57
4	455	15.3	1133	9	US-09-992-600A-57

5	453	15.3	6158	10	US-09-919-497-6	Sequence 6, Appli
6	453	15.3	6158	10	US-09-954-456-762	Sequence 762, App
7	452.5	15.2	6728	10	US-09-954-456-782	Sequence 782, App
8	452.5	15.2	6728	10	US-09-880-107-3946	Sequence 3946, Ap
9	444	15.0	5416	10	US-09-954-456-786	Sequence 786, App
10	444	15.0	5416	10	US-09-880-107-2094	Sequence 2094, Ap
11	440.5	14.8	4149	12	US-10-044-090-104	Sequence 104, App
12	440.5	14.8	5086	10	US-09-880-107-3947	Sequence 3947, Ap
13	440.5	14.8	5145	9	US-09-925-299-206	Sequence 206, App
14	440.5	14.8	5145	10	US-09-925-299-206	Sequence 206, App
15	440.5	14.8	5432	12	US-10-044-090-22	Sequence 22, Appl
16	438	14.8	2235	10	US-09-962-436-308	Sequence 308, App
17	435	14.7	3690	12	US-10-044-090-448	Sequence 448, App
18	433	14.6	2192	10	US-09-925-301-42	Sequence 42, Appl
19	432.5	14.6	1486	10	US-09-925-302-247	Sequence 247, App
20	431.5	14.5	3226	10	US-09-954-456-725	Sequence 725, App
21	426.5	14.4	4908	9	US-10-001-887-33	Sequence 33, Appl
22	425.5	14.3	2520	10	US-09-880-107-3685	Sequence 3685, Ap
23	423	14.2	481	10	US-09-815-343-1516	Sequence 1516, Ap
24	418.5	14.1	511	9	US-09-736-457-1138	Sequence 1138, Ap
25	418.5	14.1	511	9	US-09-902-941-1138	Sequence 1138, Ap
26	418.5	14.1	511	9	US-09-849-626-1138	Sequence 1138, Ap
27	415.5	14.0	2823	10	US-09-919-497-7	Sequence 7, Appli
28	414.5	14.0	2542	9	US-09-954-531-961	Sequence 961, App
29	414.5	14.0	2542	10	US-09-964-824A-255	Sequence 255, App
30	409	13.8	3394	10	US-09-880-107-2178	Sequence 2178, Ap
31	407	13.7	477	10	US-09-815-343-1517	Sequence 1517, Ap
32	397.5	13.4	2691	10	US-09-925-302-64	Sequence 64, Appl
33	394.5	13.3	1797	9	US-09-978-295A-613	Sequence 613, App
34	394.5	13.3	1797	9	US-09-978-637-613	Sequence 613, App
35	394.5	13.3	1797	9	US-09-978-192A-613	Sequence 613, App
36	394.5	13.3	1797	9	US-09-999-832A-613	Sequence 613, App
37	394.5	13.3	1797	9	US-09-978-189-613	Sequence 613, App
38	394.5	13.3	1797	9	US-10-174-590-331	Sequence 331, App
39	394.5	13.3	1797	9	US-10-176-758-331	Sequence 331, App
40	394.5	13.3	1797	9	US-10-175-737-331	Sequence 331, App
41	394.5	13.3	1797	9	US-10-173-706-331	Sequence 331, App
42	394.5	13.3	1797	9	US-10-175-738-331	Sequence 331, App
43	394.5	13.3	1797	9	US-10-175-752-331	Sequence 331, App
44	394.5	13.3	1797	9	US-10-176-482-331	Sequence 331, App
45	394.5	13.3	1797	9	US-10-176-757-331	Sequence 331, App

ALIGNMENTS

RESULT 1
US-09-745-763-198
; Sequence 198 Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallie, Edward R.
; Collins-Racine, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 198:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Alignment Scores:
Pred. No.: 9,15e-151 Length: 2930
Score: 2936.00 Matches: 545
Percent Similarity: 99.63% Conservatives: 0
Best Local Similarity: 99.63% Mismatches: 2
Query Match: 98.89% Indels: 1
DB: 10 Gaps: 0

US-09-763-712A-2 (1-547) x US-09-745-763-198 (1-2930)

QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnLeuAsnLeuThrGlnValGln 20
DB 652 ATGTATCTCATATATGTGTCATCATGAATC-AACAAGCTGAACCTGACCCAGGTGCAG 710
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
DB 711 CAGAGGAACCTCATCAGCAATCTGACGGCTGTGGATGACACAGCCAGGCTATCCAG 770
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
DB 771 CGAATCAAGAACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGGACACG 830
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
DB 831 GATTGGCTGAAGAGAGAAAGTGAGAGCTTGCAGACCTGGCTGCCAACAACCTTGCCTTG 890
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
DB 891 GCCAAGCCCAACACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 950
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
DB 951 CAGATGGAGAATCATCACCACTATCTCTCAAGCCCAAGCAGACACCTGAAAGACCTGCAG 1010
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
DB 1011 GACTTACAAAGATGCAGAGATAGAAGCAGCATCAAGTTCAACCACTGGAGGAACGC 1070
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
DB 1071 TTCCACCTCTTTCAGACGGATATGTGAACATCATTTAGCAATATCATGACAGCCAC 1130
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
DB 1131 CACCTCGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTT 1190
QY 181 ThrLysHisThrAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
DB 1191 ACCAAACACACAGATGATCTGACCTCCTTGAATAATACCTGGCCCAACATCCGTTTGGAT 1250
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
DB 1251 TCTGTTTCTCTCAGGATGCAACAGATTTGATGAGTTCAGGCTGACACTGAAGTAGCC 1310

QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlnLeuIle 240
DB 1311 AACTTATCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1370
QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGly 260
DB 1371 AAGAATTTTACAATACTACAAGGTCCACGGGGCCCGGGGGTCCAAAGAGGTGACAGAGGA 1430
QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlnLysGlyGluLysGlyGluPro 280
DB 1431 TCCAGGAGCCCTCGCCCACTGCGCAACAGGACAAAGAGAGAGAGAGGGAGCCCT 1490
QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProAlaGlyProGlyGlu 300
DB 1491 GGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCCCGGAGAG 1550
QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
DB 1551 CGTGGCGCAAAAGGATCTAAAGGCTCCAGGGCCCAAGGGCTCCCGTGGTCCCTGGG 1610
QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
DB 1611 AAGCCCGGCCCTCAGGGCCCCAGTGGGACCCAGGCCCGCCCGGGCCCAAGAGAG 1670
QY 341 GlyLeuProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
DB 1671 GGACTCCCGGCCCTCAGGGCCCTCTCGCTTCCAGGGACTTCCAGGGCCCGTGGGAG 1730
QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
DB 1731 CCTGGGTGCTTGGACTCGGGACTGCCAGGCTTGGCTGGGTACAGGCATGCCAGGC 1790
QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
DB 1791 CCCAAGGGCCCCCGGGCCCTCTGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAG 1850
QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
DB 1851 AATGAGCCCAACCCCGGACCGGAGGACAATAGCTGCCGCCCTCACTGGAAGAACTTCA 1910
QY 421 AspLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
DB 1911 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGT 1970
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
DB 1971 GAAGACAAGCTTTCACATCTTGTTTTTCATAAACACACTAGAGAGAACAGCAATGATA 2030
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
DB 2031 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTTCCAGACTCAGAGCTGAGAAAT 2090
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
DB 2091 GAATGAAGTGGCTGTGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCG 2150
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleValAlaGly 520
DB 2151 GATACTGGGTTCATGGCCATGGCCAGGAGAGAGACTGCTGGTGGTTGATTTATGCTGG 2210
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
DB 2211 CAGTGGAGAGATTTCCAATGTGAAGACGTCATAAATCTTTCATTTTCGAAAAAGACAGGAG 2270
QY 541 ThrValLeuSerSerAlaLeu 547
DB 2271 ACAGTACTGTCTATCTCAITTA 2291

RESULT 2

US-09-745-763-2274
; Sequence 2274, Application US/09954456
; Patent No. US20020115057A1

```

; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2274
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-2274

Alignment Scores:
Pred. No.: 4,31e-18 Length: 1410
Score: 474.50 Matches: 111
Percent Similarity: 49.67% Conservative: 41
Best Local Similarity: 36.27% Mismatches: 132
Query Match: 15.98% Indels: 23
DB: 10 Gaps: 6

US-09-763-712A-2 (1-547) x US-09-954-456-2274 (1-1410)

QY 246 LeuGlnGlyProGlyProArgGlyAspArgGlySerGlnGlyProPro 265
DB 400 ATGCTTGGCAAGCTGGCCAGTGGCCCAAGGGGACAATGGCTCTGTGGGAACCT 459
QY 266 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyProGlyPro 285
DB 460 GGACCAAAAGGGAGACACTGGGCCAAGTGGACCTCCAGGACCTCCCGTGTGCTGCCA 519
QY 286 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGly 305
DB 520 GCTGAAGAGAAAGTCCCTCGGGAGGAGGAGGACATAGGACCTCAGGCAAGCCAGGC 579
QY 306 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 325
DB 580 CCAAAAGGAGAAAGCTGGGCCCAAGGAGAAAGTAGGTGCCCAAGGATCGAGGGCTCGGCA 639
QY 326 GlyProSerGlyAspProGlyProGlyPro-----ProGlyLysGluGlyLeu 342
DB 640 GGGGCAAGAGGCTCGCAGGCCCTTAAGGAGAGCGAGGTGTCCTGGTGAGCGTGGAGTC 699
QY 343 ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly 362
DB 700 CCTGGAAGCGCAGGGGCGACGAGGGTCTGCTGGAGCCATGGGTCCCGCAGGGAAGTCCAGGT 759
QY 363 ValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys 382
DB 760 GCCAGGGGACCCCGGGATTGAAGGGGGACAAAGGACATTCCTGGAGACAAAGGCAAG 819

```

```

QY 383 GlyProProGlyProGlyProSerGlyAlaValValProLeu---AlaLeuGlnAsn 401
DB 820 GGAGAAAGTGGGCTTCCAGATGTTGCTCTCTGAGGACGACGAGTTTGAGCCCTTACAGGCA 879
QY 402 GluProThrProAlaProGluAspAsnGlyCys-----GATGGCGGGTCCAG 412
DB 880 CAAAGT-ACAGCACCTCCA-----GGCTGCTTCTCTCAGTATATAAGAAAGTTGAGCT 929
QY 413 ProProHisTrp-LysAsnPheThrAspLysCysTyTrPheSerValGluLysGlu1 432
DB 930 CTTCCCAAAATGCCAAAGTGTGGGGAGAGATTTTCAAGACGACGAGGCTTTGTAACACC 989
QY 432 ePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnTh 452
DB 990 ATTACGGAGGACAGCTGCTGTGCACACAGGCTGGTGACAGATTGGCCCTCTCCACGCTC 1049
QY 452 rArgGluGlnGlnTrpIleLysLysGlnMetValGlyArg---GluSerHisTrp1 471
DB 1050 TGCCGCTGAGAATGCCGCTTGCACACAGGCTGCTGCTAGTAAAGACGAGGCTGCTTCT 1109
QY 471 eGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAs 491
DB 1110 GAGCATGACTGATCCAAAGACAGAGGGGCAAGTTACCTACCCACAGGAGAGTCCCTGCT 1169
QY 491 pTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyG1 511
DB 1170 CTATTCCAACCTGGGCCCGGAGGAGCCCAACGAT-----GATGGCGGGTCCAG 1217
QY 511 uAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAs 531
DB 1218 GGACTGTGTGGAGATCTTCCCAATGGCAAGTGGAAATGACAGGGCTTGTGGAGAAAGCG 1277
QY 531 nAsnPheIleCysGlu 536
DB 1278 TCTTGTGTCTCGAG 1293

RESULT 3
US-09-924-340-57
; Sequence 57, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91 US2 REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Patent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
US-09-924-340-57

```

Alignment Scores:
Pred. No.: 3,99e-17 Length: 1133
Score: 455.00 Matches: 98
Percent Similarity: 47.46% Conservative: 14
Best Local Similarity: 41.53% Mismatches: 50
Query Match: 15.33% Indels: 74
DB: Gaps: 7

US-09-763-712A-2 (1-547) x US-09-924-340-57 (1-1133)

QY	246	LeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	265
DB	140	CTCCAGGCCCCAGGCCAGCGGGTCCAGAGGAGGAGGAGGCCCAAGGTAACTCC	199
QY	266	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro	285
DB	200	GGTGAAGGCGACACAGGATTCAGAGCCAGCCAGGCTTCCGGGCCACCCGGGTCCC	259
QY	286	AlaGlyGluArgGlyProLleGlyProAlaGlyProGlyProGlyGluArgGlyGly	305
DB	260	CCTGGATTCACGAGCAAGTGTGATCATCCTGGCCACCTGGCCCTCAAGCAGAGAAGGC	319
QY	306	SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro	324
DB	320	AGCGAAGGATTCAGAGCCCATCAGGCCTGCCTGGCTCCCTGGGCCACCCGGGACCTCT	379
QY	325	-----GlnGlyProSer-----	328
DB	380	GGGATTACAGGGCCCCCGCGGTCTGGATGTTTGGATGGGAAGGATGGCAAGCCTGGCTTTG	439
QY	329	---GlyAspProGlyProGlyProPro---	337
DB	440	AGGGGGACCTCGGTCTCTGCTGGCCCCCTGGACTCATGGGACACCGGGCTTTAAGGGG	499
QY	338	-----GlyLysGluGlyLeuProGlyProGln-----	349
DB	500	AAACAGGACATCCTGCCTCCAGGACCTAAGGTGACTGTGGCAACACCGAGTCTCTCT	559
QY	350	GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu	369
DB	560	GGCAGCACTGGCGGCGCTGGCGAGAGGGTGAACCTGTGTCCATGGACCCAGGGAAGA	619
QY	370	ProGlyLeuPro-----	373
DB	620	CCCGTCCCCGGGACACGTGGTGGCCACAGGCGCTCCAGGCCAGCCAGGACCAAGCTGGG	679
QY	374	-----	374
DB	680	ATCTCTGCATGGGTCTGAAGAGAGACGAGGAGCCACCGAGAAAGGGGCTTGCAGGC	739
QY	375	ValProGlyMetProGlyProLysGlyProProGlyProPro-----	392
DB	740	CTCCAGGCGGAGCGGCGGCCCCAGGTCCTCAAGGTCCTCCAGGCTATGGCAAGATGGGT	799
QY	393	AlaValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaPro	407
DB	800	GCAACAGGACCATGGGCCAGCAGGACATCCTGGCATCCTGGGCC 847	

RESULT 4
US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715

; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782

Alignment Scores:
Pred. No.: 2.77e-16 Length: 6728
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 15.24% Indels: 49
DB: 10 Gaps: 8

US-09-763-712A-2 (1-547) x US-09-954-456-782 (1-6728)

Qy 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 265
Db 1485 GTTCAAGGACCCCTGGCCCTGCTGGAGAGGAAGAAAGCGAGGAGCTCGAGGTGAACCC 1544
Qy 266 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro----- 280
Db 1545 GGACCCACTGGCTGCCCGGACCCCTGGAGGCTGCTCCCGGTGAAGCTGGTCTGCCT 1604
Qy 281 -----GlyProGlyProAlaGlyGluArgGlyProIleGly 293
Db 1605 CTGCGCGCAGATGGTGTGCTGCCAAGGCTCCCGTGGTGAACGTGGTCTCCTGCTGC 1664
Qy 294 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 310
Db 1665 CCCGCTGCCCCCAAGGATCTCTGCTGAAGCTGCTGCCCGGTGAAGCTGGTCTGCCT 1724
Qy 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyThrValGlyGluPro 330
Db 1725 GGTGCCAAGGTCTGACTGGAAGCCCTGGCAGCCCTGGCTGCTGATGGCAAACT----- 1778
Qy 350 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 361
Db 1836 GCCCGCTGCAGGTGCTGATGGATTCCTCGGACCCCTGGCTGCTGGAGAGCC 1895
Qy 376 ProGlyMetProGlyProGlySerProGlyProGlyProGlyProSerGly----- 392
Db 1956 GATGAGAGGTGGAGCTCAGGACCCCTGGCTGCTGCTGCCCGGTGGAGAGAGGT 2015
Qy 393 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 406

Db 2016 GAACAAGGCCCTGGTGGCTCCCC---GGATTCCAGGGTCTCCCTGGTCTGCTGCTCCT 2072
Qy 407 -----ProGluAspAsnGlyCysPro 413
Db 2073 CCAGGTGAAGCAGCAACCTGGTGAACACAGGGGTGTTCT 2111
RESULT 8
US-09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3946
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274615
US-09-880-107-3946

Alignment Scores:

Pred. No.: 2.77e-16 Length: 6728
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 15.24% Indels: 49
DB: 10 Gaps: 8

US-09-763-712A-2 (1-547) x US-09-880-107-3946 (1-6728)

Qy 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 265
Db 1485 GTTCAAGGACCCCTGGCCCTGCTGGAGAGGAAGAAAGCGAGGAGCTCGAGGTGAACCC 1544
Qy 266 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro----- 280
Db 1545 GGACCCACTGGCTGCCCGGACCCCTGGAGGCTGCTCCCGGTGAAGCTGGTCTGCCT 1604
Qy 281 -----GlyProGlyProAlaGlyGluArgGlyProIleGly 293
Db 1605 CTGCGCGCAGATGGTGTGCTGCCAAGGCTCCCGTGGTGAACGTGGTCTCCTGCTGC 1664
Qy 294 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 310
Db 1665 CCCGCTGCCCCCAAGGATCTCTGCTGAAGCTGCTGCCCGGTGAAGCTGGTCTGCCT 1724
Qy 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 330
Db 1725 GGTGCCAAGGTCTGACTGGAAGCCCTGGCAGCCCTGGCTGCTGATGGCAAACT----- 1778
Qy 331 ProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProPro--- 349
Db 1779 ---GGCCCCCTGGTCTCCCGCGGTCAAGATGGTCCCGGACCCCTCGAGGCCACCTGCT 1835
Qy 350 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 361
Db 1836 GCCCGTGTGAGGTGCTGATGGATTCCTCGACCTAAAGGTGCTGCTGGAGAGCC 1895
Qy 362 GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal----- 375


```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2094
; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03464
US-09-880-107-2094

Alignment Scores:
Pred. No.: 6,53e-16 Length: 5416
Score: 444.00 Matches: 115
Percent Similarity: 43.40% Conservative: 23
Best Local Similarity: 36.16% Mismatches: 79
Query Match: 14.95% Indels: 102
DB: 10 Gaps: 10

US-09-763-712A-2 (1-547) x US-09-880-107-2094 (1-5416)
QY 197 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAsp 216
Db 2280 GTCCATTGGAAGCGAGGTCTCTTGGACCCCGGCTGATGGAACAAGGGTGAAC 2339
QY 217 ThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer----- 234
Db 2340 CTGGTGTGGTGTGTGGCAGCTGTGGTCCATCTGTGTAGTGACTCCAGGAG 2399
QY 235 -----LysHisGlyGlnLeuIleLysAsn 242
Db 2400 AGAGGGGTGCTGCTGCATACCTGTGAGGCAAGGAGAAAGGTGAACCTGTCTCAGAG 2459
QY 243 -----PheThrIleLeuGlnGlyPro 249
Db 2460 GTGAATTTGGTAACCTGSCAGAGATGGTGTCTGTGCTCATGTGTCTGTA-GGTGCC 2518
QY 250 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProGly 266
Db 2519 CTGTGCTCTGTGCGGACACAGGTACCGGGCGGAGGTGGGCTGCTGTCTGTGTGT 2578
QY 267 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 286
Db 2579 CTGTGTGTCTCGGGGAAGCCCTGTGTAACGTGGGAGGTGCTGTCTGTGCCCCAAC 2638
QY 287 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 303
Db 2639 GCATTGTGCTGCGGCTGTGCTGTCTGTCAACCGGGTGTAAAGGAGAAAGAGAGGC 2698
QY 304 LysGlySerLysGlySerGln----- 310
Db 2699 AAAGGGCTTAAGGGTGAACACGGTGTGTGTGTCCACAGGCCCGGTTGGAGCTGTGTGC 2758
QY 311 -----GlyProLysGlySerArgGlySer----- 318
Db 2759 CCAGCTGTGTCACAAATGGTCCCGCGGTCTGTGTGAAGTGTGGTATGAGAGGCCCGCT 2818
QY 319 -----ProGlyLysProGlyProGlnGlyPro 327
Db 2819 GGTATGACTGGTTTCCCTGTGTCTGTGACGAGTGGTCCCGACGACCTCTGTGTAT 2878
QY 328 SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln--- 346
Db 2879 TCTGGCCCTCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 2938
QY 347 -----GlyProProGlyPheGln 352
Db 2939 GACCAAGTTCAGTTGGCGAAGTAGTGATGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2998
QY 353 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 366
Db 2999 GTGAGAGGGTCCCTCTGTGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3058
QY 367 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 386
Db 3679 GTGAGAGGGTCCCTCTGTGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3728
```

```
Db 3059 CAGGCTCTTTTGTGTCTCTCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3118
QY 387 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 403
Db 3119 CTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3178
QY 404 -----ThrProAlaProGluAspAsnGlyCysPro 413
Db 3179 GGGGCGCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3232

RESULT 11
US-10-044-090-104
; Sequence 104, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 104
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.19
; NAME/KEY: unsure
; LOCATION: 1492-1495, 1663, 1669, 4012, 4055
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-104

Alignment Scores:
Pred. No.: 7.92e-16 Length: 4149
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 14.84% Indels: 53
DB: 12 Gaps: 6

US-09-763-712A-2 (1-547) x US-10-044-090-104 (1-4149)
QY 248 GlyProProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 267
Db 1307 GTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1366
QY 268 Thr-----Gly 269
Db 1367 AATGGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1426
QY 270 AsnLysGlyGlnLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 289
Db 1427 GCCAAAGGCCCTTAAGGGTGAACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1486
QY 290 GlyProIleGlyProAlaGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 309
Db 1487 GGCNNNNNGTCCAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1546
QY 310 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro----- 327
Db 1547 CCTGTATGACTGGTTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1606
QY 328 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 346
Db 1607 ATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1666
QY 347 -----GlyProProGlyPhe 351
Db 1667 GNGACCAAGGACACAGGAGAGTAGGAGAGTAGGAGAGTAGGAGAGTAGGAGAGTAG 1726
QY 352 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 365
```

```
Db 1727 GCTGGTGAAGGGTCCCTCTCTGGAGAGGCTGGTACTGCTGGACCTCCTGGCACTCCAGGT 1786
QY 366 ProArgLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 385
Db 1787 CTCAGGGTCTTCTGCTGCTCTGCTATCTGGGCTCTCCCTGGCTCGAGAGGTGAACGT 1846
QY 386 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 402
Db 1847 GGTCTACCTGGTGTGCTGCTGGTGTGGAACCTGGTCTCTTGGCATTCGCGCCCT 1906
QY 403 Pro-----ThrProAlaProGluAsnGlyCysPro 413
Db 1907 CCTGGGCGCGTCTCTCTGCTGTGTGGTGTAGTCTCTGGAGTCAACGGTGTCTCT 1963

RESULT 12
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Alignment Scores:
Pred. No.: 9,53e-16 Length: 5086
Score: 440.50 Matches: 114
Percent Similarity: 43.08% Conservative: 23
Best Local Similarity: 35.85% Mismatches: 80
Query Match: 14.84% Indels: 102
DB: 10 Gaps: 10

US-09-763-712A-2 (1-547) x US-09-880-107-3947 (1-5086)
QY 197 IleArgLeuAspSerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAsp 216
Db 1950 GTCCATTGTGAAGCGGAGGTCTCTTGGACCCCGGCTGTATGAAACAAAGGGTGAAC 2009
QY 217 ThrGluValAlaAsnSerValIleMetGluMetLysLeuValAspSer----- 234
Db 2010 CTGGTGTGGTGTGCTGGGCACTGCTGTCCATCTGTGTCTAGTGGACTCCCGAGGAG 2069
QY 235 -----LysHisGlyGlnLeuIleLysAsn 242
Db 2070 AGAGGGTGTCTCTGGCATACTCTGGAGCAAGGGAAGAGGGTGAACCTGTCTCAGAG 2129
QY 243 -----pheThrIleLeuGlnGlyPro 249
Db 2130 GTCAAAATTGGTAACCTGGCAGAGATGFGCTCGTGGTCTCATGCTGTGTA-GGTGCC 2188
QY 250 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProGly 266
```

```
Db 2189 CCTGTCTCTCTCTGGAGCCACAGGTGACCGGGCGGAAGCTGGGGCTGCTGTCTCTGTGGT 2248
QY 267 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAla 286
Db 2249 CCTGTGTGTCTCTGGGGAAGCCCTGTGTGAACGTGGCGAGGTCTGTCTGTGGCCCAAC 2308
QY 287 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 303
Db 2309 GGATTGTGTGTCTGGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2368
QY 304 LysGlySerLysGlySerGln----- 310
Db 2369 AAAGGCGCTTAAGGCTGAACAAACGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2428
QY 311 -----GlyProLysGlySerArgGlySer----- 318
Db 2429 CCAGCTGTGTCAAAATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2488
QY 319 -----ProGlyLysProGlyProGlnGlyPro 327
Db 2489 GGTATGACTGTTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2548
QY 328 SerGlyAspProGlyProProGlyProGlyLysGlyGluLeuProGlyProGln----- 346
Db 2549 TCTGGCCCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2608
QY 347 -----GlyProProGlyPheGln 352
Db 2609 GACCAAGGTCCAGTGTGGCGAACTGGAGAAGTAGGTGTGTGTGTGTGTGTGTGTGTGTGT 2668
QY 353 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 366
Db 2669 GGTGAAGGGTCCCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2728
QY 367 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 386
Db 2729 CAGGCTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2788
QY 387 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 403
Db 2789 CTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2848
QY 404 -----ThrProAlaProGluAsnGlyCysPro 413
Db 2849 GGGGCGCGTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2902

RESULT 13
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
```

```
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.: 9.63e-16 Length: 5145
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 14.84% Indels: 53
DB: 9 Gaps: 6

US-09-763-712A-2 (1-547) x US-09-925-299-206 (1-5145)
Qy 248 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 267
Db 2248 GGTCTGTGTGTCCTCGGGGAAGCCCTGTGAACGTGTGAGTGGTCTCTGCTGCC 2307
Qy 268 Thr-----Gly 269
Db 2308 AATGATTTGTGTCCTGCTGCTGCTGCTCAACCTGTGCTAAAGGAGAAAGAGGA 2367
Qy 270 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyPro 289
Db 2368 GCCAAGGGCCCTAGGCTGAAACGGTGTGTTGGTCCACAGGCCCTGTGGAGCTGCT 2427
Qy 290 GlyProLleGlyProAlaGlyProProGlyGluArgGlyGlySerLysGlySer 309
Db 2428 GGCCAGCTGTCTCAATAGTGTCCCGGCTGCTGGAAGTCGTGTGATGGAGGCC 2487
Qy 310 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 327
Db 2488 CTGTGTATGACTGTTTCCCTGGTCTGCTGGAGAGGCTGCTGCTGCAACCTGCT 2547
Qy 328 ---SerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGln 346
Db 2548 ATTTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2607
Qy 347 Pro-----ThrProAlaProGluAspAsnGlyCysPro 413
Db 2848 CTTGGGCGCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2904

RESULT 14
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US2002005562A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
```

```
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.: 9.63e-16 Length: 5145
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 14.84% Indels: 53
DB: 10 Gaps: 6

US-09-763-712A-2 (1-547) x US-09-925-299-206 (1-5145)
Qy 248 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 267
Db 2248 GGTCTGTGTGTCCTCGGGGAAGCCCTGTGAACGTGTGAGTGGTCTCTGCTGCC 2307
Qy 268 Thr-----Gly 269
Db 2308 AATGATTTGTGTCCTGCTGCTGCTGCTCAACCTGTGCTAAAGGAGAAAGAGGA 2367
Qy 270 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyPro 289
Db 2368 GCCAAGGGCCCTAGGCTGAAACGGTGTGTTGGTCCACAGGCCCTGTGGAGCTGCT 2427
Qy 290 GlyProLleGlyProAlaGlyProProGlyGluArgGlyGlySerLysGlySer 309
Db 2428 GGCCAGCTGTCTCAATAGTGTCCCGGCTGCTGGAAGTCGTGTGATGGAGGCC 2487
Qy 310 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 327
Db 2488 CTGTGTATGACTGTTTCCCTGGTCTGCTGGAGAGGCTGCTGCTGCAACCTGCT 2547
Qy 328 ---SerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGln 346
Db 2548 ATTTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2607
Qy 347 Pro-----GlyProProGlyPhe 351
Db 2608 GGTGACCAAGTCTCAGTTGGCCGAAGTAGGTGAGTGGTGGTCCCTGCTGCTTC 2667
Qy 352 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 365
Db 2668 GCTGTGAGAGGGTCCCTCTGGAGAGGCTGTTACTGCTGGACCTCTCCGACTCC 2727
Qy 366 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGlyProGln 385
Db 2728 CCTCAGGCTCTCTTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2787
Qy 386 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 402
Db 2788 GGTCTACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2847
```

```
QY 403 Pro-----ThrProAlaProGluAspAsnGlyCysPro 413
|||
Db 2848 CCTGGGGCCCGTGGTCTCTCTGCTGTGTGGGTAGTCCTGGAGTCAACGGTCTCT 2904
|||
RESULT 15
US-10-044-090-22
; Sequence 22, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 5432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
US-10-044-090-22

Alignment Scores:
Pred. No.: 1,01e-15 Length: 5432
Score: 440.50 Matches: 114
Percent Similarity: 43.08% Conservative: 23
Best Local Similarity: 35.85% Mismatches: 80
Query Match: 14.84% Indels: 102
DB: 12 Gaps: 10

US-09-763-712A-2 (1-547) x US-10-044-090-22 (1-5432)
QY 197 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAsp 216
|||
Db 2283 GTCTATTGGAAGCAGAGTCTCTTGGACCCCGGCTGTATGGAACAAAGGTTGAAC 2342
|||
QY 217 ThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer----- 234
|||
Db 2343 CTGGTGTGGTGTGGTGGGACATGCTGCTCCATCTGCTAGTGACCTCCACAGGAG 2402
|||
QY 235 -----LysHisGlyGlnLeuIleLysAsn 242
|||
Db 2403 AGAGGGTGTCTGTGCATACCTGGAGGCAAGGAGAAAGGTGAACCTGTGCTCAGAG 2462
|||
QY 243 -----PheThrIleLeuGlnGlyPro 249
|||
Db 2463 GTGAATTGGTAACCTGGCAGAGATGGTGTGGTGTCTCTGTGTCTGTGA-GGTGCC 2521
|||
QY 250 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProProGly 266
|||
Db 2522 CCTGGTCTCTGTGGACCCACAGGTGACCGGGCGAAGCTGGGGCTGCTGCTGCTGCT 2581
|||
QY 267 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 286
|||
Db 2582 CCTGTGTGCTCTCGGGAAGCCCTGTGAACGTGTGTGAGTGTGCTGCTGCTGCCCAAT 2641
|||
QY 287 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 303
|||
Db 2642 GGATTTGTCTGTGCTGCTGTGCTCAACCTGTGTGCTTAAGAGAGAGAGAGAGGCC 2701
|||
QY 304 LysGlySerLysGlySerGln----- 310
|||
Db 2702 AAAGGCGCTAAGGCTGAACAAACGGTGTGTGTGTCCACAGGCCCGCTTGGAGCTGCTGCC 2761
|||
QY 311 -----GlyProLysGlySerArgGlySer----- 318
|||
Db 2762 CCAGCTGGTCCAATGGTCCCCCGGCTCTGTGTGAAGTGTGTGTGTGAGAGGCCCGCT 2821
|||
QY 319 -----ProGlyLysProGlyProGlnGlyPro 327
|||
```

```
Db 2822 GGTATGACTGGTTTCCCTGGTGTCTGTGGACGACTGGTCCCGCCAGGACCTCTGTGTA 2881
|||
QY 328 SerGlyAspProGlyProProGlyProProGlyLysGlyGluProGlyProGln--- 346
|||
Db 2882 TCTGGCCCTCTCTGGTCCCGCTCTGTGTGGGAAAGAGGGCTTCTGTGCTCTGTGCT 2941
|||
QY 347 -----GlyProProGlyPheGln 352
|||
Db 2942 GACCAAGTCCAGTTGGCCGAACGTGAGAGTAGGTGCAGTTGGTCCCGCTGGCTCGCT 3001
|||
QY 353 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 366
|||
Db 3002 GGTGAGAGGGTCCCTCTCTGGAGAGGGTGTGTGTGCTGTGGACCTCTCTGGCCTCC 3061
|||
QY 367 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGlyProGlyPro 386
|||
Db 3062 CAGGGTCTTCTTGGTGTCTCTGTGTATCTGGGTCTCCCTGGCTCGAGAGGTGAACGTG 3121
|||
QY 387 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 403
|||
Db 3122 CTACCAGGTGTGCTGTGCTGTGGGTGAACCTGGTCTCTTGGCATTGCCGCCCTCT 3181
|||
QY 404 -----ThrProAlaProGluAspAsnGlyCysPro 413
|||
Db 3182 GGGGCCCGTGGTCTCTCTGTGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 3235
|||
```

Search completed: March 21, 2003, 12:31:42
Job time : 295.779 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:08:41 ; Search time 1808.03 Seconds
(without alignments)
4899.781 Million cell updates/sec

Title: US-09-763-712A-2
Perfect score: 2969
Sequence: 1 MYSHNVIMNLNLTLQVQ.....EDVNNFICKDRETVLSSAL 547

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlp
-Q=/cgn2_1/USPRO_pool/US09763712/runat_14032003_100950_18109/app_query.fasta_1.1877
-DB=EST -QFMT=fastap -SUFMT=first -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pt0 -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -CGN_1.1.4575 -runat_14032003_100950_18109 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_esti.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2765	93.1	3305	11	BC009162	BC009162 Mus muscu
2	1460.5	49.2	957	14	BQ955927	BQ955927 AGENCOURT
3	1394.5	47.0	936	14	BQ891432	BQ891432 AGENCOURT
4	1256	42.3	861	14	BQ713873	BQ713873 AGENCOURT
5	1251	42.1	906	13	BI456109	BI456109 603172765
6	1219.5	41.1	884	14	BQ934501	BQ934501 AGENCOURT
7	1022	34.4	580	12	BE910803	BE910803 601661855
8	1007	33.9	668	14	BQ674807	BQ674807 AGENCOURT
9	957	32.2	601	10	BE290299	BE290299 601089246
c 10	916.5	30.9	967	9	AL568743	AL568743 AL568743
11	901	30.3	581	14	BQ127513	BQ127513 1160h06.y
12	842	28.4	552	10	AW958053	AW958053 EST370123
13	831	28.0	638	10	BB248064	BB248064 BB248064
14	777.5	26.2	511	12	BF805750	BF805750 QV1-CI017
15	767	25.8	808	14	BQ771366	BQ771366 UI-M-F10-
16	669	22.5	500	14	BM676508	BM676508 UI-E-EJ0-
17	663	22.3	507	14	BM713891	BM713891 UI-E-EJ0-
c 18	650	21.9	410	12	BF856923	BF856923 QV1-FT020
19	627.5	21.1	591	13	BI445884	BI445884 dai33d10.
20	602	20.3	1082	14	BM907108	BM907108 AGENCOURT
21	600	20.2	357	10	AW379436	AW379436 CM4-HT024
22	594	20.0	350	9	AA012704	AA012704 RPUL402CG
23	588	19.8	339	9	AA304251	AA304251 EST17006
c 24	560	18.9	517	9	AL543000	AL543000 AL543000
25	537	18.1	310	9	AA361740	AA361740 EST71069
26	516	17.4	465	10	AW240221	AW240221 up30b07.y
27	503.5	17.0	839	13	BI697412	BI697412 603348154
28	478	16.1	943	14	BQ887163	BQ887163 AGENCOURT
c 29	477	16.1	679	14	W27610	W27610 35b6 Human
30	460	15.5	441	10	BB749743	BB749743 BB749743
31	458.5	15.4	569	10	AV616076	AV616076 AV616076
32	457.5	15.4	1029	17	CNS04VYG	AL309897 Tetraodon
33	456.5	15.4	792	13	BI456015	BI456015 603170460
34	456	15.4	420	12	BE929813	BE929813 RC3-GN004
35	455.5	15.3	896	14	BQ955616	BQ955616 AGENCOURT
36	453.5	15.3	877	12	BG681943	BG681943 602629884
37	452	15.2	688	13	BI249235	BI249235 602995511
38	452	15.2	3951	11	BC013626	BC013626 Mus muscu
39	451.5	15.2	921	14	BQ922102	BQ922102 AGENCOURT
40	451.5	15.2	1192	14	BM905999	BM905999 AGENCOURT
41	449.5	15.1	907	14	BQ222937	BQ222937 AGENCOURT
42	448	15.1	946	14	BQ942234	BQ942234 AGENCOURT
43	447	15.1	862	12	BG171919	BG171919 602333020
44	446.5	15.0	608	13	BM425818	BM425818 pgf2c.Bk0
45	446	15.0	651	10	BB613976	BB613976 BB613976

ALIGNMENTS

RESULT 1
BC009162
LOCUS BC009162 3305 bp mRNA linear HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to scavenger receptor with C-type lectin,
clone IMAGE:2811487, mRNA.
ACCESSION BC009162.1 GI:14714370
VERSION BC009162
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg,R.
TITLE Direct Submission


```

QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2121 GAATGGAAGTGCAGCGGTACCTGTGATTACAAAAAAGTGAAGAGTGGACAACA 2180

QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2181 GATAACTGGGCGAGTGGCCATGGCCAGGAGAAAGTGTGCTGGCTGATTACGAGGA 2240

QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2241 CAGTGGATGATCTCCAGTGTGATGAATCAATACTTCATTTGTGTGAGAAAGGAG 2300

QY 541 ThrValLeuSerAlaLeu 547
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2301 GCAGTACCATCATCCATATTA 2321

RESULT 2
BQ955927 957 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8763247 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313500
DEFINITION 5', mRNA sequence.
ACCESSION BQ955927
VERSION BQ955927.1 GI:22371405
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13740 row: 1 column: 13
High quality sequence stop: 665.
Location/Qualifiers
1. 957
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC library."
BASE COUNT 260 a 269 c 258 g 169 t . 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.61e-91 Length: 957
Score: 1460.50 Matches: 284
Percent Similarity: 93.73% Conservative: 15
Best Local Similarity: 89.03% Mismatches: 13
Query Match: 49.19% Indels: 8
DB: 14 Gaps: 3

US-09-763-712A-2 (1-547) x BQ955927 (1-957)

QY 92 AsnSerGln---LeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 110
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 8 AATTCGGGGATCTCAGCTCATTCACAGTGCAGATGGACAACATACCACTATCTCACAG 67

QY 111 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 130

```

```

Db 68 GCCAACGACGACGAGCCTGAAAGACCTTCAGGACTTTACACAAGGATACAGAAAATAGAACA 127
QY 131 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGlnThrAspIleValAsn 150
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 138 GCTGTCAAGTTACAGCAACTTGAGGAAGCTTCCAGGCTCTTTGAGACAGATATTGTGAAC 187
QY 151 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 170
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 188 ATCATTAGCAACATCAGCTACACAGCCCATCCTGAGGACACTGCCAGCAATCTGAAT 247
QY 171 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSerLeu 190
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 248 GATGTTAGGACCACATGCACAGACACTTGACCAGACACACGATGACCTGACCTCCTTG 307
QY 191 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 210
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 308 AATAACACACTAGTCAACATCCGCTGGATTCTATTTCTCAGGATGAGCAAGACATG 367
QY 211 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 230
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 368 ATGAGTCAAAAGTTAGACACTGAAGTGGCCAACTTATCAGTGGTTATGAAGAGATGAAA 427
QY 231 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 250
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 428 CTGGTTGACTCCAAGCAGCGTCACTCAAGAAGTATTTACCATTTCTACAAGGTCCTCCT 487
QY 251 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 270
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 488 GCCCCAGAGGTCCAAAAGGTGACAGAGATCTCAGGAGCACCACTGGTCCCACTGGCAAC 547
QY 271 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProGlyGluArgGly 290
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 548 AAGGACAGAAAGGAGAGAGAGAGAGAGCGCTGCTCCACCTGGCCCTGCGGGTGAGAGGGC 607
QY 291 ProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySerGln 310
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 608 ACAATTTGACCACTGCGCCCTCTCGAGAGCGTGGCAGCAAGGATCCAAAGGCTCACAG 667
QY 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlyProGlyProGlyAsp 330
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 668 GGTCCCAAAGGATCTGCTGGGTCCCCCAGGAGAGCGTGGCCCTCAAGAGACCTAGTGGGGAC 727
QY 331 ProGlyProGlyProGlyProGlyLysGlyGluLeuProGlyProGlyProGlyProGly 350
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 728 CCAGGACCACTGCTCCAGGAGAGAGAGAGAGAGTGGACTCCCTGGCCCTCANGGCCCTCCTGGC 787
QY 351 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu-Pr 370
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 788 TTGAG-GGACTACAGGGCACTGTGGGTGAGCGCTGGAGTACCTGGACCTCGGGGGTTGCC 846
QY 370 oGlyLeuProGlyValProGlyMetPro--GlyProLysGlyProPro-GlyProProGly 389
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 847 AGCTTGGCCAGGGGTGCCAGGATGCCCTGGGGCCCTAAGGGACCACTGGGCCCTCCAAG 906
QY 389 yPro---SerGlyAlaValValProLeuAla---LeuGlnAsnGluPro 403
    |||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 907 GCCCCTCAGGAGACAAGTGGAGCCACTTGGCCCTCCTGCAGAAATGAACCA 955

RESULT 3
BQ891432 936 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT_8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
DEFINITION 5', mRNA sequence.
ACCESSION BQ891432
VERSION BQ891432.1 GI:22283446
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 936)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

```

TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.

CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM13790 row: n column: 19
High quality sequence start: 25
High quality sequence stop: 632.

FEATURES

source

1..936
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6332754"
/lab_host="NIH_MGC_130"
/note="Organ: otocysts; Vector: pCMV-SPORT6.1.ccdB;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."

BASE COUNT 273 a 250 c 244 g 163 t 6 others
ORIGIN

Alignment Scores:

Pred. No.: 5,71e-87 Length: 936
Score: 1394.50 Matches: 272
Percent Similarity: 95.05% Conservative: 16
Best Local Similarity: 89.77% Mismatches: 9
Query Match: 46.97% Indels: 6
DB: 14 Gaps: 1

US-09-763-712a-2 (1-547) x BQ891432 (1-936)

QY 60 ThrAspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAla 79
|||||
Db 38 ACCGATTGGCTTAAGGAAGATACAGAGCTTCGACACATTGGCTGCCAACAACTCTGCC 97
QY 80 LeuAlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThr 99
|||||
Db 98 CTGGCCAAAGCCCAATGACACCTAGAGGATATGAATAGCCAGCTCAGCTCATTCACA 157
QY 100 GlyGlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeu 119
|||||
Db 158 GGTCACTGGACACAACTATACCACTATCTCACAGGCCAACGACAGAGCCTGAAAGACCTT 217
QY 120 GlnAspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGlu 139
|||||
Db 218 CAGACTTACAGAGGATACAGAAATAGAACAGCTGTCAAGTTCAGCCAACTTGAGGAA 277
QY 140 ArgPheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyThrAla 159
|||||
Db 278 CGCTTCAGGCTCTTCAGACAGATATGTGAACATCATTTAGCAACATCAGCTACAGCC 337
QY 160 HisHisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThr 179
|||||
Db 338 CATCACTGAGGACACTGACCAACATCTGAATGTGTAGGACCACATGACAGACACC 397
QY 180 LeuThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAlaAsnIleArgLeu 199
|||||
Db 398 TTGACCAGACACAGGATGACCTGACCTCTTGAATTAACACATAGTCAACATCCGCTTG 457
QY 200 AspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluVal 219
|||||
Db 458 GATTCTATTCTCTCAGGATGACGCAAGACATCATGAGCTCAAGTTAGACACTGAAGTG 517
QY 220 AlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeu 239

Db 518 GCCAACTTATCACTGGTGTATGGAAGAGATGAAACTGGTTGACTCCAGCAGCGTCAGCTC 577
QY 240 IleLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArg 259
Db 578 ATCAAGAACTTTACCACTTACAAAGTCTCTCGGCCCCAGAGGTCCAAAGAGTGACAGA 637
QY 260 GlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGlu 279
Db 638 GGATCTCAGGACCACTGTCTCACTGCAACAGAGGACAGAAAGAGAGAGAGAGAG 697
QY 280 ProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGly 299
Db 698 CTTGGTCCACTGGCTCGGGTGAGAGGGGCACAATTGGACCACTCGGCCCTCCTGGA 757
QY 300 GluArgGlyGlyLysGlySerLysGlySerGlnGly-ProLysGlySerArgGlySer-P 319
Db 758 GAGCGTGGCAGCAAGGATCCAAAGGCTCACAGGGNTCCAAAGGGATCTCGTGGGTCCC 817
QY 319 roGlyLysProGlyProGlnGlyProSerGlyAspProGly-ProProGlyProProGly 338
Db 818 CAGGGAAGCTGGCCCTCAAGGAGCTAGTGGGACCCAGGACCCAGGTCACACAGGN 877
QY 339 -LysGluGlyLeuPro-GlyProGlnGlyProPro---GlyPheGlnGlyLeuGlnGly 356
Db 878 CAGGGATGAGACTCCCTGGGCCCTCAGGGNCCNTCCTGGGCTCNCAGGGGACTACAAGG 936

RESULT 4

BQ713873

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BQ713873 861 bp mRNA linear EST 16-JUL-2002
AGENCOURT_8482828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240
5', mRNA sequence.

BQ713873 GI:21852772

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

CDNA Library Preparation: ResGen, Invitrogen Corp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM13721 row: n column: 01

High quality sequence stop: 598.

FEATURES

source

1..861
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6306240"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.ccdB; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."

BASE COUNT 221 a 253 c 239 g 147 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.94e-77 Length: 861

Score: 1256.00 Matches: 229

Percent Similarity: 91.01% Conservative: 14

RESULT 5	BI456109	603172765F1	NCI_CGAP_Mam5	906 bp	linear	EST 21-AUG-2001
LOCUS	BI456109					
DEFINITION	Mus musculus cdna clone IMAGE:5251888 5', mRNA sequence.					
ACCESSION	BI456109					
VERSION	BI456109.1					
KEYWORDS	GI:15246765					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					

```

Db 423 TCCAAAGGCTCAGAGGTCCTCAAGAGTCTCGTGGTCCCGAGTGAAGCTGGCCCTCAA 482
QY GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 345
Db 483 GGACCTAGTGGGACCCAGGACCCAGGTCACACGAGGTCACAGGAGGATGAGTCCCTGGCCCT 542
QY GlnGlyProGlyPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 365
Db 543 CAGGGCCCTCCTGGCTTCCAGGAGCTACAGGACATGTGGGTGAGCTGGAGTACCTGGA 602
QY ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGlyProGlyProPro 385
Db 603 CCTCGGGGTTGCCAGGCTTCCAGGGTCCAGGAGTCCAGGATGCTGGGCTTAAGGACCACT - 661
QY GlyProGlyProSerGlyAlaValValProLeuLeuAlaLeuGlnAsnGluProThrPro 405
Db 662 AGCCCTCCAGGCCCCTCAGGAGATG - GAGCCCATTTGCTGAGAAATGAACCAAGCCCA 720
QY AlaProGlu - AspAsnGlyCysProProHisTrpLysAsnPheThrAsp - --LysCysT 424
Db 721 GCATCAGAGGTTCAACGCGATGTCCTCTCACTGGAAGAACTTTCCACGGATTATGGCT 780
QY yTrpPheSerVal - --GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspL 443
Db 781 ACTATTATTTTCATTGCACAAAGAAATTTTGAAGATTGCTTAAGCGTTCTGTGGAGAAA 840
QY ySerSerHis 446
Db 841 AATTTTCCAAT 851

```

```

RESULT 6
LOCUS BQ934501
DEFINITION AGENCOURT_8765685 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330794
5' mRNA sequence.
ACCESSION BQ934501
VERSION BQ934501.1 GI:22349884
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.

```

```

CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13785 row: m column: 03
High quality sequence stop: 352.
FEATURES
Location/Qualifiers
1..884
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6330794"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"

```

```

note="Organ: otocysts; vector: pCMV-SPORT6.1.cdb;
Site:1: EcoRV; Site:2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 282 a 225 c 205 g 172 t
ORIGIN

```

```

Alignment Scores: 6,64e-75 Length: 884
Pred. No.: 1219,50 Matches: 255
Score: 92,54% Conservativeness: 18
Percent Similarity: 86,44% Mismatches: 14
Best Local Similarity: 41,07% Indels: 9
Query Match: 14 Gaps: 4
DB:
US-09-763-712a-2 (1-547) x BQ934501 (1-884)
QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 7 ATGTATTCTCAGAGCGTGTTCATCATGAACCTCAACAACTGAACCTTAACCCAGGTTCCAG 66
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 67 CAGAGGAACCTTATCTCAAAATCTGCAGCAGTCTGTGGATGACACAGCCTGGCCATCCAG 126
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 127 CCAATTAAGAATGATTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACACC 186
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 187 GATTGGCTAAAGGAAAAAGTACAGAGCTTGCAGACATTGGCTGCCAACAACTTGCCTCG 246
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThr-Gl 100
Db 247 GCCAAAGCCCAACAAATGACACCTAGAGGATATGAATAGCCAGCTCAGCTCATTTACAGG 306
QY 100 yGlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGl 120
Db 307 TCAGATGGACAACATTACCACTATCTCACAGGCCAAGAGAGAGCCTGAAGACCTTCA 366
QY 120 nAspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluAla 140
Db 367 GGACTTACACAAGGATACAGAAATAGAACAGCTGTCTCAAGTTTCAGCCAACTTGAGGAG 426
QY 140 qPheGlnLeuPheGluThrAspIleValIleSerAsnIleSerAsnIleSerTyrThrAlaH 160
Db 427 CTTCCAGGCTTTTGAGACAGATATTGTGAACATCATATTGACATCATCATCATCATCAT 486
QY 160 shisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLe 180
Db 487 TCACCTGAGGACACTGACCAAGCAATCTGAATGATGTTAGGACACATGACACACACCTT 546
QY 180 uThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAs 200
Db 547 GACCAGACACAGGATGACCTGACCTCTTGAATAC-ACACTAGTCAACATCCGCTTGA 605
QY 200 pSerValSerLeuArgMetGlnAsnLeuMetArgSerArgLeuAspThrGluValAl 220
Db 606 TTTCTATTCTCTCAGATGTCACAGACATGATGAGGTCAAGTTAGACACTGAAGTGGC 565
QY 220 aAsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuI 240
Db 666 CAACCTTATCAGTGGTTATGGAAGAGATGAACTGGTTGACTCCAAGCAGGTCAGCTCAT 725
QY 240 elysAsnPheThrIleLeuGlnGlyProGlyProArgGly-ProArgGlyAspArgG 260
Db 726 CAAGAACCTTTACCTTCTACAAGGTCTCTCTGGCCCCAGAGGTCAAAAGGGGACAGAG 785
QY 260 lysSerGlnGlyProPro---GlyProThr-GlyAsnLysGlyGlnLys-----GlyGlu 276
Db 786 GGATCTCAGGGGACCACTGGGTCCCACTTGCACACAGGGGGAACAAAAGGGAGAGAA 845
QY 277 LysGlyGluPro---GlyProPro---GlyProAlaGly 287
Db 846 AGGGGAGAGAGCCTGTGGTCCACCTGGGGCCCTGGGGCC 884
RESULT 7
BE910803

```

```

LOCUS       BE910803               580 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION  601661855F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962292 5',
            mRNA sequence.
ACCESSION   BE910803
VERSION     BE910803.1   GI:10407765
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 580)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM9128 row: m column: 13
            High quality sequence stop: 580.
            Location/Qualifiers
                1..580
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:3962292"
                /clone_lib="NCI_CGAP_Mam1"
                /tissue_type="tumor, biopsy sample"
                /dev_stage="10 months, virgin"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigator
                providing samples: Gilbert Smith, NIH"

BASE COUNT  159 a 153 c 161 g 107 t
ORIGIN
Alignment Scores:
Pred. No.:      1.9e-61      Length:      580
Score:          1022.00      Matches:    177
Percent Similarity: 96.34%      Conservative: 7
Best Local Similarity: 92.67%      Mismatches: 0
Query Match:    34.42%      Indels:     0
DB:             12          Gaps:        0

US-09-763-712A-2 (1-547) x BE910803 (1-580)

QY  304  LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGly 323
      |||||||
Db   6  AAAGGATCCAAAGGTCACAGGGTCCCAAGGATCTCGTGGTCCCGAGGGAAGCTGGC 65
      |||||||
QY  324  ProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyGluGlyLeuPro 343
      |||||||
Db   66  CTTCAAGGACCTAGTGGGGACCCAGGACCCAGGATCCACAGGCAAGATGATCCCT 125
      |||||||
QY  344  GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
      |||||||
Db   126  GGCCCTCAGGGGCCCTCTGGCTTCCAGGGACTACAGGGCAGCTGTGGTGAGCCTGGAGTA 185
      |||||||
QY  364  ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
      |||||||
Db   186  CTTGACCTCGGGGGTTCGACGGCTTGCAGGGGTGCCAGGGATGCTGGGCTTAAGGA 245
      |||||||
QY  384  ProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 403
      |||||||
Db   246  CCACCTGGCCCTCCAGGCCCTCCAGAGCAATGGAGCAATGGCTCTGCAGATGAACCA 305
      |||||||
QY  404  ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCys 423

```

```

Db  306  ACCCCAGCATCAGAGGTCAACGGATGTCGCTCACTGGAAGAACTTCACAGATAATGC 365
      ||||||| ||| |||||||
QY  424  TyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLys 443
      |||||||:|||||
Db  366  TACTATTTTTCATTGGAAGAAATTTTGAAGATGCTAAGCTTTCTGTGAAGACAA 425
      |||||||
QY  444  SerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMet 463
      |||||||:|||||
Db  426  TCTTCCCTCTCGTTTTCATAAATCAAGAGAAGAACAGCAATGGATAAAAGCATACC 485
      |||||||
QY  464  ValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 483
      |||||||:|||||
Db  486  GTGGGAGAGAAGCAATGGATCGGCTCACAGACTCAGAACAGGAAGCAATGAAG 545
      |||||||
QY  484  TrpLeuAspGlyThrSerProAspTrpLysAsn 494
      |||||||
Db  546  TGGCTAGACGGGTACCTGTTGATTACAAAAC 578
      |||||||

RESULT 8
BO674807
LOCUS       AGENCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672
DEFINITION  5', mRNA sequence.
ACCESSION   BO674807
VERSION     BO674807.1   GI:21785641
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 668)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM2377 row: c column: 09
            High quality sequence stop: 656.
            Location/Qualifiers
                1..668
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6212672"
                /clone_lib="NIH_MGC_102"
                /tissue_type="epidermoid carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed
                by Ling Hong in the laboratory of Gerald M. Rubin
                (University of California, Berkeley) using ZAP-CDNA
                synthesis kit (Stratagene) and Superscript II RT (Life
                Technologies). Note: this is a NIH_MGC Library."

BASE COUNT  216 a 148 c 167 g 137 t
ORIGIN
Alignment Scores:
Pred. No.:      2.4e-60      Length:      668
Score:          1007.00      Matches:    174
Percent Similarity: 99.43%      Conservative: 0
Best Local Similarity: 99.43%      Mismatches: 1
Query Match:    33.92%      Indels:     0
DB:             14          Gaps:        0

```


ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 967)
 AUTHORS Li, W. B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1..967
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DE005YH04"
 /clone_lib="LPL_FL002_PL1"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
 ORIGIN

Alignment Scores:
 Pred. No.: 6.28e-54 Length: 967
 Score: 916.50 Matches: 174
 Percent Similarity: 91.15% Conservative: 1
 Best Local Similarity: 90.62% Mismatches: 14
 Query Match: 30.87% Indels: 5
 DB: 9 Gaps: 3

US-09-763-712A-2 (1-547) x AL568743 (1-967)

QY 356 GlyThrValGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyVal 375
 Db 947 GGCACCGTGGAGCGCTGGGTCT---GGACCTCGGAGCTGCAG---CTTGCCCTGGGTA 894

QY 376 ProGlyMetProGlyProGlyProGlyProGlyProGlyProSerGlyAlaValVal 395
 Db 893 CGAGCATGCCAGSCCCCAAGG---CCCCCCCCCTCTCTGC---CCATCAGGCGGTTGTG 838

QY 396 ProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProProHis 415
 Db 837 CCCCTGSCCTGCACAAATGAGCAACCCCGGACCGGAGGACAATGGCTGCCCGCTCAC 778

QY 416 TrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAsp 435
 Db 777 TGGAAAGACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGAT 718

QY 436 AlaLysLeuPheCysLysSerHisLeuValPheIleAsnThrArgGluGlu 455
 Db 717 GCAAAGCTTTCTGTGAGACAAAGTCTCACATCTGTTTTCATAAACACTAGAGAGAA 658

QY 456 GlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAsp 475
 Db 657 CAGCAATGGATAAAAAACAGATGCTAGGAGAGAGAGCCACTGGATCGCCCTCACAGAC 598

QY 476 SerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrp 495
 Db 597 TCAGAGCGCTGAAATCAATGGAAGTGGTGGATGGGACATCTCCAGACTACAAAAATGG 538

QY 496 LysAlaGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 515
 Db 537 AAAGCTGGACACCCGGATAACTGGGGTCTATGC-CATGGGCCAGGAGAACTGTCTCTGG 479

QY 516 LeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCys 535
 Db 478 TTGATTATGCTGGCAGTGGAGCAATTTCCCAATGTGAAGACGTCAATAACTTCATTTC 419

QY 536 GluLysAspArgGluThrValLeuSerSerAlaLeu 547
 Db 418 GAAAAAGACAGGAGACAGACTGTGCATCTGCATTGA 383

RESULT 11
 BQ127513
 LOCUS BQ127513 581 bp mRNA linear EST 19-APR-2002
 DEFINITION i160h06.y1 Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus musculus cDNA clone IMAGE:5946107 5' similar to TR:Q9UM15 Q9UM15 CSR2. ;, mRNA sequence.

ACCESSION BQ127513
 VERSION BQ127513.1 GI:20201424
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 581)
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarisvili, R., Williams, T., Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Other ESTs: i160h06.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center This clone is available royalty-free through LBNL; please contact the IMAGE consortium (info@image.llnl.gov) for further information
 Seq primer: -40RP from Gibco
 High quality sequence stop: 446.

FEATURES
 source
 1..581
 /organism="Mus musculus"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone="IMAGE:5946107"
 /clone_lib="Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1"
 /sex="Both for embryonic & newborn, male for adult and adult islet"
 /dev_stages="Embryonic day 10.5, E12.5, E16.5, newborn, adult, mixed"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite

chromatography and used to make this library."

```
BASE COUNT      189 a 159 c 114 g 119 t
ORIGIN

Alignment Scores:
Pred. No.:      4,37e-53      Length:      581
Score:          901.00      Matches:      177
Percent Similarity: 97.41%      Conservative: 11
Best Local Similarity: 91.71%      Mismatches: 5
Query Match:      30.35%      Indels:      0
DB:              14      Gaps:      0

US-09-763-712A-2 (1-547) x B0127513 (1-581)
QY 12 AsnAsnLeuAsnLeuThrGlnValGlnGlnArgAsnLeuLeuThrAsnLeuGlnArgSer 31
DB 3 AACAACTGTACCTACCCAGGTTTCAGCAGAGAGAACTTATCTCAAATCTGCAGCAGTCT 62
QY 32 ValAspAspThrSerGlnAlaIleGlnArgIleLysAsnAspPheGlnAsnLeuGlnGln 51
DB 63 GTGGATGACACAAAGCCTGGCCATCCAGCGAATTAAGAATGATTTCCAAATCTGCAGCAG 122
QY 52 ValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGlnSerLeuGln 71
DB 123 GTTTCCTTCAGCCCAAGAGGACACCGATTGGCTTAAGCAAAAGATACAGAGCTTGCAG 182
QY 72 ThrLeuAlaAlaAsnSerAlaLeuAlaLysAlaAsnAspThrLeuGluAspMet 91
DB 183 ACATTGGCTGCCAACAACTCTGCCCTGGCCAAAGCCAAACATGACACCCCTAGAGGATATG 242
QY 92 AsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGlnAla 111
DB 243 AATAGCAGCTCAGCTCAATTCAGAGTCAGATGGCAACATACCACCTATCTCACAGGCC 302
QY 112 AsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThrAla 131
DB 303 AAGCAGCAGAGCCTGNAAGACCTTCAGGACTTACACAGAGGTACAGAAATAGACAGCT 362
QY 132 IleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsnIle 151
DB 363 GTCAAGTTCAGCCAACTTCAGGAACGCTTCCAGGCTTTTGAGACAGATATCTGAACTC 422
QY 152 IleSerAsnIleSerThrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsnGlu 171
DB 423 ATTAGCAACATCAGTACACAGCCCATCACCTGAGGACACTGACCAGCAATCTGAAATGAT 482
QY 172 ValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeuAsn 191
DB 483 GTTAGGACCACTGACAGACACCTTGACCAGACACACCGATGACCTGACCTCTTGAAT 542
QY 192 AsnThrLeuAlaAsnIleArgLeuAspSerValSerLeu 204
DB 543 AACACACTAGTCAACATCGCTTGGATTCTATTCTCTC 581

RESULT 12
AW958053
LOCUS          AW958053          552 bp      mRNA      linear      EST 01-JUN-2000
DEFINITION    EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION     AW958053
VERSION       AW958053.1 GI:8147736
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 552)
AUTHORS       Heyde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt
               I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
               Quackenbush,J.
TITLE         Assessment of gene expression patterns in a model of colon tumor
               metastasis using a 19,200 element cDNA microarray
JOURNAL       Unpublished (2000)
COMMENT       Contact: John Quackenbush
```

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 115
Seq primer: Reverse.

FEATURES
source
Location/Qualifiers
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"

BASE COUNT 176 a 112 c 143 g 121 t
ORIGIN

Alignment Scores: 4,95e-49 Length: 552
Pred. No.: 842.00 Matches: 149
Score: 98.68% Conservative: 1
Percent Similarity: 98.03% Mismatches: 1
Best Local Similarity: 28.36% Indels: 1
Query Match: 10 Gaps: 0
DB: 0

US-09-763-712A-2 (1-547) x AW958053 (1-552)

```
QY 397 LeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrp 416
DB 1 CTGGCCCTGCAGAAATGAGCAACCCCGCCGAGGAGCAATGGCTGCCGCTCACTCG 60
QY 417 LysAsnPheThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAla 436
DB 61 AAGAAGCTTCACAGACAAATGCTACTATTTTTCAGTTTGAGAAAGAAATTTTGGAGTGA 120
QY 437 LysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGln 456
DB 121 AAGCTTTTCTGTGAGACAAAGTCTTCACATCTGTGTTTTCATAAACACTAGAGGAGAACAG 180
QY 457 GlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSer 476
DB 181 CAATGGATAAAAAACAGATGCTAGGAGAGAGAGAGCCACTGGCTCGGCTCACAGACTCA 240
QY 477 GluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLys 496
DB 241 GAGCGTGAATAATGAATGGAAGTGGCTGGATGGAGACATCTCCAGACTACAAAAATTGAAA 300
QY 497 AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 516
DB 301 GCTGACACCGCCGATAACTGGGGTTCATGGCCATGGGCCAGGAGAGACTGTGCTGGGTG 360
QY 517 IleTyAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhe-IleCysGln 536
DB 361 ATTTATGCTGGCAGGGGAACGATTTCCAATCTGAAGACGTCAATAACTTAATTATTCGA 420
QY 536 uLysAspArgGluThrValLeuSerSerAlaLeu 547
DB 421 AAAAGACAGGGAGACAGTACTGGCATCTGCATTA 454
```

RESULT 13
BB248064
LOCUS BB248064 638 bp mRNA linear EST 23-OCT-2001
DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus
ACCESSION BB248064
VERSION BB248064.2 GI:16355610
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 638)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

BQ771366 808 bp mRNA linear EST 26-JUL-2000
 UI-N-F10-byu-g-09-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
 IMAGE:5702432 5', mRNA sequence.

eukaryota; Metazoa; Chordata; Sciurognathi; Muridae; Murinae; Mus.
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 808)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

(BMAP)
Seq primer: pYX-5.

```
seq primer: p1a.3.
Location/Qualifiers
1..808
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702432"
```

```
/clone_lib="NIH_BMAP_F10"
/tissue type="whole brain"
```

```
/dev_stage="embryo 12.5dpc"  
/lab_host="DH10B (T1 phage resistant)"
```

```

/iao_host= Dn10B (IT phage resistant)
/note="Organ: Brain; Vector: pYX- Asc; Site_1: Ecor I;
site 2: Not I. The library was constructed according

```

Site 2; not 1, the library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was

primer containing a NotI site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoRI adaptor, digested with NotI and then cloned

then took 1 µg adapter/digested with NotI and then ligated directionally into pYX-Asc vector. The library tag sequence located between the NotI site and the polyA tail

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the CAGCCACGAC. This library was created for the University of Iowa and is not available outside the project.

Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,

program coordinator.		number of female teachers (n=100)		number of female students (n=100)	
		a	b	c	d
270	a	158	c	173	q
				206	t
					1 others

0
0
0
0
0
4

0
0
0
4

n
0
2
4

0
0
0
4

.

4

2
0

```
cores:
  1.11e-43      Length: 808
```

Matches:	131
Conservative:	9

Conservative:	10
Mismatches:	0
Indels:	0
Similarity:	25.83%
Identity:	87.33%

```

#1000000
14
Gaps: 0

```

2A-2 (1-547) x BQ771366 (1-808)

aLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrpLys 417

TCTGCAATGAACCAACCCAGCATCAGAGGTCAACGGATGTCCGCTCACTGGAAG 61

nPhThrAspLysCysTyrPheSerValGluLysGluIlePheGluAspAlaLys 437

.....


```
Db 62 AACTTCACAGATAAATGCTACTATTTCATTGGAAAAAGAAATTTTGAAGATGCTAAG 121
Qy 438 LeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGln 457
    |||||
Db 122 CTTTTCGTGAGACAATCTCCCATCTCGTTTTTCATAACTCAAGAGAGAACAGCAA 181
    |||||
Qy 458 TrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGlu 477
    |||||
Db 182 TGGATAAAAAAGCATACCGTGGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGAA 241
    |||||
Qy 478 ArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAla 497
    |||||
Db 242 CAGGAAGCGAATGGAAGTGGCTAGACGGGTCACTGTGTGATTACAAAACCTGGAAAGCT 301
    |||||
Qy 498 GlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIle 517
    |||||
Db 302 GGACAACCAGATAACTGGGGCAGTGGCCATGGCCAGAGAGACTGTCTGGCTTGATT 361
    |||||
Qy 518 TyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLys 537
    |||||
Db 362 TACGAGGACAGTGGGAATGACTTCCAGTGTGATGAATCAATAACTTCATTGTGAGAAG 421
    |||||
Qy 538 AspArgGluThrValLeuSerSerAlaLeu 547
    |||||
Db 422 GAAGGGAGGAGGAGTACCATCATCATATTA 451
    |||||
```

Search completed: March 21, 2003, 12:20:20
Job time : 1823.03 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 06:54:47 ; Search time 2261.11 Seconds
(without alignments)
4401.885 Million cell updates/sec

Title: US-09-763-712A-2_COPY_206_547

Perfect score: 1940

Sequence: 1 MOODLMRSRLTEVANLSVI.....EDVNNFICKDRETVLSSAL 342

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spcol/US09763712/runat_14032003.100949.18101/app_query.fasta_1.1877
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn.1.1.6828_erunat_14032003.100949.18101 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1940	100.0	2983	9	AB005145	AB005145 Homo sapi
2	1934	99.7	2005	6	AX454442	AX454442 Sequence
3	1934	99.7	2005	6	AX490920	AX490920 Sequence
4	1934	99.7	2641	6	AX047353	AX047353 Sequence
5	1934	99.7	3058	9	AB038518	AB038518 Homo sapi
6	1932	99.6	1886	9	HSB802985	AL713657 Homo sapi
7	1801	92.8	2637	10	AB078434	AB078434 Mus muscu
8	1795	92.5	3291	10	AB038519	AB038519 Mus muscu
9	1664	85.8	4330	9	AB052103	AB052103 Homo sapi
10	1021.5	52.7	130763	2	AC112416	AC112416 Rattus no
11	981	50.6	182029	2	AC114677	AC114677 Mus muscu
12	981	50.6	193208	2	AC102618	AC102618 Mus muscu
13	954.5	49.2	169088	2	AC016128	AC016128 Homo sapi
14	954.5	49.2	188439	9	AP000915	AP000915 Homo sapi
15	935.5	48.2	71044	2	AC024368	AC024368 Homo sapi
16	916.5	47.2	187635	2	AP001022	AP001022 Homo sapi
17	914.5	47.1	178022	2	AP000900	AP000900 Homo sapi
18	842	43.4	188255	2	AP000939	AP000939 Homo sapi
19	551	28.4	2215	10	BC026446	BC026446 Mus muscu
20	540	27.8	3636	9	AB007829	AB007829 Homo sapi
21	540	27.8	3685	6	E32511	E32511 Scavenger r
22	540	27.8	3810	6	E32509	E32509 Scavenger r
23	491.5	25.3	1265	10	RATSPD	M81231 Rat palmona
24	483.5	24.9	3508	3	AF053538	AF053538 Alvinella
25	481.5	24.8	1183	10	BC003705	BC003705 Mus muscu
26	481.5	24.8	1253	10	MUSSPD	L40156 Mus musculu
27	474.5	24.5	1410	6	AX334792	AX334792 Sequence
28	474.5	24.5	1410	9	HSMRNAPD	X65018 H.sapiens m
29	474	24.4	1454	4	BTLSPD	X75911 B.taurus mR
30	473.5	24.4	1301	9	BC022318	BC022318 Homo sapi
31	473	24.4	1284	4	AF509590	AF509590 Bos tauru
32	471	24.3	2377	9	BC008760	BC008760 Homo sapi
33	471	24.3	5676	6	E07265	E07265 cDNA encodi
34	471	24.3	5676	9	HUMCALV	D90279 Human mRNA
35	471	24.3	7138	9	HUMPALV	M76729 Human pro-a
36	465	24.0	6114	10	CRUPALV	M76730 Chinese ham
37	462.5	23.8	4990	10	CRUCOLVIFA	L08863 Cricetus
38	462	23.8	2777	5	GGU07973	U07973 Gallus gall
39	462	23.8	4428	6	AX146422	AX146422 Sequence
40	462	23.8	4428	6	AX146424	AX146424 Sequence
41	462	23.8	5575	5	AF137273	AF137273 Gallus ga
42	461	23.8	5551	10	AF272662	AF272662 Rattus no
43	461	23.8	6156	10	AB009993	AB009993 Mus muscu
44	459.5	23.7	810	6	AR014116	AR014116 Sequence
45	459	23.7	5650	5	AB015440	AB015440 Rana cate

ALIGNMENTS

RESULT 1

AB005145 2983 bp mRNA linear PRI 21-NOV-2001
LOCUS Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
DEFINITION
AB005145
ACCESSION
VERSION
AB005145.1 GI:17026100
KEYWORDS
SOURCE
ORGANISM
Homo sapiens female tissue_lib:placenta cDNA to mRNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuo,H., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.
The membrane-type collectin CL-P1 is a scavenger receptor on
vascular endothelial cells
J. Biol. Chem. 276 (47), 44222-44228 (2001)
21570232
2 (bases 1 to 2983)
REFERENCE
Ohtani,K.
Direct Submission
Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail: ohtaniasahikawa-med.ac.jp,
Tel:+81-166-68-2393, Fax:+81-166-68-2399)
FEATURES
Location/Qualifiers
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
1..2983
/gene="CL-P1"
71..2299
/codon_start=1
/gene="CL-P1"
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDDFAEEVEVQSFYKRGFQIEGTQCTKCKNNWALKFSTILLY
ILCALLTIVAILGYKVKEDNDVNGMETSQTYDDKLTAVESDLKLGDOTGKKA
SNSELSFRSDILROOLREITEKTSKNKDTLEKLOASGDALVDROSOLKTELNN
SLFITVVKLTLYNGYVNLQODTSVLOGNLQOMYSNVIMNINLNLTVQVQRN
LITNQRSDDYTSQAIRIKNQFLQKQVFLQAKDITDLWKEKVSQSLQTLAANSALA
KANNDTLEDMSOLANSFTGMENITTSOANEONLKDQLDHLKDAENRTAIFKNLEE
RQLPETDITVNTISNISTYTAHLRLTSLNNEVRTDITLTFKDDTLTSLNLTANI
RLDSVLSRQDLMSRLDTEVANLSVMEEMKLVDSKHGLIKNFTILOGPPGPRGP
RGDSGSPGPGTGNKGKGEKGEPPGPGAGERPIGPAGPPGERGKSGKSGPK
GSRGSPGPGPGSDGPPGPGKEGLPGQPGPGQGLGTGVEGPGVPGPLPG
LPGVPMFGPGKPGPGPGSGAVPLALONEPTPAPDNGCPPHWKNFTDKCYFSVE
KEIFEDAKLFCEDKSSHLVFINTRREQQWIKQOMYGRESHWITGLTDSRENEMKWL
TSPDYKNMKAGQPDNWNHGHGPGEDCAGLIYAGQWNDFQEDVNNFICEKDRETVLSS
AL"
BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN
Alignment Scores:
Pred. No.: 1,95e-58 Length: 2983
Score: 1940.00 Matches: 342
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Indels: 0
Gaps: 0
DB: 9
US-09-763-712a-2_COPY_206_547 (1-342) x AB005145 (1-2983)
QY 1 MetGlnGluAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1271 ATGCAACAGATTGTAGAGTTCAGGCTAGACACTGAAGTAGCACTATCAGTGATT 1330
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 1331 ATTTGAAAGAAATGAAGCTAGTAGACTCCAGACATGGTCAGCTCATCAAGAAATTTTACAATA 1390

QY 41 LeuGlnGlyProProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
|||||
Db 1391 CTACAAGGTCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 1450
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyPro 80
|||||
Db 1451 GGCCCCAACTGGCAACAGGACACAAAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1510
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGly 100
|||||
Db 1511 GCGGGTGAGAGAGGCCCAATTGGACCACTGTGTCCCGGAGAGGCTGGCGGCAAGGA 1570
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
|||||
Db 1571 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 1630
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
|||||
Db 1631 GGCCCCAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGCCCT 1690
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||
Db 1691 CAGGGCCCTCCTGGCTTCAGGGACTTCAGGGACCTTCAGGGAGGCTGGGGTGCCTGGA 1750
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||
Db 1751 CCTCGGGGACTGCCAGCTTGCCTGGGTACCAGGCATGCCAGGCCCAAGGCCCCCCC 1810
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||
Db 1811 GGGCCCTCCTGGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCAACCCCG 1870
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
|||||
Db 1871 GCACCGGAGGACAACTGGCTGCCGCCCTCACTGGAAGAACTTCACAGCAAAATGCTACTAT 1930
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
|||||
Db 1931 TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAAGACAAAGCTTCA 1990
QY 241 HisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1991 CATCTTGTGTTTCAATAACACTAGAGAGGAAACAACAATGGATAAAAAACAGATGGTAGGG 2050
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||
Db 2051 AGAGAGGCCACTGGATCGCCTCACAGACTCAGAGCGTGAAAATGAATGGAAGTGGCTG 2110
QY 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||
Db 2111 GATGGGACATCTCCAGACTACAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCAT 2170
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||
Db 2171 GGCCATGGGCCAGGAGAAAGACTGTGCTGGTGTGATTATGCTGGGAGTGGAAAGATTTC 2230
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
|||||
Db 2231 CAATGTGAAGACGTCATCACTTCACTTGGGAAAAAGACAGGGAGACAGTACTGTCTATCT 2290
QY 341 AlaLeu 342
|||||
Db 2291 GCATTA 2296
RESULT 2
LOCUS AX454442
DEFINITION Sequence 27 from Patent WO0208284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Baker, K.P., Ferrara, N., Gerber, H., Gerttsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gerttsen, Mary E. (US) ; Goddard,
Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
I. (US)
FEATURES Location/Qualifiers
source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 2,38e-58 Length: 2005
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservatives: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 6 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AX454442 (1-2005)
QY 1 MetGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 478 ATGCAACAAGATTGTATGAGTTCGAGGTGTAGACACTGAAGTAACTATATCAGTGATT 537
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLysAsnPheThrIle 40
Db 538 ATGGAGAATAAGCTAGCTAGCTCAAGCATGTCACCTCATCANGAATTTTACAATA 597
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 598 CTACAAGTCCACCGGCCCCAGGGTCCAGAGGTGACAGAGGATCCCGAGGACCCCT 657
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 658 GSCCCCAACTGGGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProAlaGlyProGlyGluArgGlyGly 100
Db 718 GCGGTGACAGAGGCCCAATTCAGCAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGA 777
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
Db 778 TCTAAGGCTCCACGGGCCCAAGGCTCCCGTGTTCCTCCCTGGGAAGCCCGCCCTCAG 837
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyProGlyGlyLeuProGly 140
Db 838 GSCCCCACTGGGAGCCACAGGCCCGCCCGCCACAGGCAAGAGAGAGAGAGAGAGAG 897
QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 898 CAGGGCCCTCTGGCTTCCAGGACTTCAGGACACCGTTGGGAGCGCTGGGTGCTGGA 957
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyPro 180
Db 958 CCTGGGGAGCTGCCAGGCTTGCCTGGGTACCCAGGATGCCAGGCCCAAGGCCCGCC 1017
QY 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||

Db 1018 GGCCCTCTCGGCCCATCAGAGGGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCG 1077
QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
Db 1078 GCACCGGAGGACAATAGCTGCCCGCTACTGGAAGAACTTCACACACAAATCTACTAT 1137
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 240
Db 1138 TTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGAAGCAAGCTTCA 1197
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1198 CATCTGTTTTCATAAACAACACTAGAGAGAACAGCAATGATATAAAAAACAGATGGTAGG 1257
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1258 AGAGAGAGCCACTGGATCGCCCTCAGACTCAGAGCGTGAAAATGAATGGAAGTGGCTG 1317
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1318 GATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCGGATAAATCTGGGTCAT 1377
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1378 GGCCATGGGCCAGGAGAGACTGTGCTGGTTCATTTATGCTGGGAGTGGACGATTC 1437
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 1438 CAATGTGAAGAGCTCANTACTTCATTCGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
QY 341 AlaLeu 342
Db 1498 GCATTA 1503
RESULT 3
AX490920
LOCUS AX490920 Sequence 27 from Patent WO200690.
DEFINITION AX490920
ACCESSION AX490920
VERSION AX490920.1 GI:22323797
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
AUTHORS Baker, K.P., Ferrara, N., Gerber, H., Gerttsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0200690-A 27 03-JAN-2002;
Genentech, Inc. (US)
FEATURES Location/Qualifiers
source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 2,38e-58 Length: 2005
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservatives: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 6 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AX490920 (1-2005)
QY 1 MetGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||

Db	478	ATGCACACAGATTGATGAGGTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT	537
QY	21	MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIle	40
Db	538	ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAATTTTACAATA	597
QY	41	LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	60
Db	598	CTACAAGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCT	657
QY	61	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro	80
Db	658	GGCCCCAACTGGCAACAGGAGACAAAGAGAGAGAGGGGAGCTGGACACCTGGCCCT	717
QY	81	AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly	100
Db	718	GCGGGTGAGAGAGGCCCAATTTGACCAAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGA	777
QY	101	SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln	120
Db	778	TCTAAAGGCTCCCGAGGGCCCAAGGCTCCCGTGGTTCCTCGGAGGCGCGCCCTCAG	837
QY	121	GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro	140
Db	838	GGCCCCAGTGGGACCCAGCCCGCCCGCCCAACAGGAGACTTCCCGCGCCCT	897
QY	141	GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly	160
Db	898	CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCGTGGGGTGCCTGA	957
QY	161	ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro	180
Db	958	CCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGATGCCAGGCCCAAGGCGCCCT	1017
QY	181	GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro	200
Db	1018	GGCCCTCTGGCCCATCAGAGGGGTGGTGGCCCTGGCCCTGCAGATGAGCCAAACCCG	1077
QY	201	AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr	220
Db	1078	GCACCGGAGGACAAATAGTGGCGGCTCACTGCAAGAACTTCACACAAATGCTACTAT	1137
QY	221	PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer	240
Db	1138	TTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCA	1197
QY	241	HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly	260
Db	1198	CATCTTGTTCATAAACACTAGAGAGGACACCAATGGATAAAAAACAGATGGTAGGG	1257
QY	261	ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu	280
Db	1258	AGAGAGGCGACTGGATCGGCTCACAGACTCACAGCGTGAAATCAATGGAAGTGGCTG	1317
QY	281	AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis	300
Db	1318	GATGGGACACTCTCCAGACTACAAAAATGGAAAGCTGGACAGCGGATAACTGGGTCAT	1377
QY	301	GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTrpAlaGlyGlnTrpAsnAspPhe	320
Db	1378	GGCCATGGGCGCAGAGAGACTGTGCTGGGTGATTTATGCTGGGAGTGGAGCAATTC	1437
QY	321	GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer	340
Db	1438	CAATGTGAAGACGTCAATAACTTCAATTCGAAAAAGACAGGGAGACACTGTGTCTCT	1497
QY	341	AlaLeu	342
Db	1498	GCATTA	1503
RESULT	4		
AX047353			
LOCUS	AX047353	2641 bp	DNA linear PAT 15-DEC-2000

DEFINITION	Sequence 39 from Patent WO0068380.		
ACCESSION	AX047353		
VERSION	AX047353.1		
KEYWORDS	GI:11876599		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Bandman, O., Hillman, J.L., Tang, Y.T., Lal, P., Yue, H., Baughn, M.R., Lu, D.A. and Azimzai, Y.		
TITLE	Extracellular matrix and adhesion-associated proteins		
JOURNAL	Patent: WO 0068380-A 39 16-NOV-2000;		
FEATURES	Incyte Genomics, Inc. (US)		
Source	Location/Qualifiers		
	1..2641		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/note="incyte ID No: 4586187CB1"		
BASE COUNT	823 a	628 c	608 g
ORIGIN	582 t		
Alignment Scores:			
Pred. No.:	2,87e-58	Length:	2641
Score:	1934.00	Matches:	341
Percent Similarity:	99.71%	Conservative:	0
Best Local Similarity:	99.71%	Mismatches:	1
Query Match:	99.69%	Indels:	0
DB:	6	Gaps:	0
US-09-763-712a-2_COPY_206_547 (1-342) x AX047353 (1-2641)			
QY	1	MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle	20
Db	912	ATGCACACAGATTGATGAGGTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT	971
QY	21	MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIle	40
Db	972	ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAATTTTACAATA	1031
QY	41	LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	60
Db	1032	CTACAAGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT	1091
QY	61	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro	80
Db	1092	GGCCCCAACTGGCAACAGGAGACAAAGGAGAGAGGGGAGCTGGACACCTGGCCCT	1151
QY	81	AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly	100
Db	1152	GCGGGTGAGAGAGGCCCAATTTGACCAAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGA	1211
QY	101	SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln	120
Db	1212	TCTAAAGGCTCCCGAGGGCCCAAGGCTCCCGTGGTTCCTCCCGGAGGCGCGCCCTCAG	1271
QY	121	GlyProSerGlyAspProGlyProProGlyProProGlyLysGlyGluGlyLeuProGlyPro	140
Db	1272	GGCCCCAGTGGGACCCAGGCCCGCCCGCCCAAGGAGGAGGACTCCCGCCCT	1331
QY	141	GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly	160
Db	1332	CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCGTGGGGTGCCTGA	1391
QY	161	ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro	180
Db	1392	CCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGATGCCAGGCCCAAGGCGCCCT	1451
QY	181	GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro	200
Db	1452	GGCCCTCTGGCCCATCAGAGGGGTGGTGGCCCTGGCCCTGCAGATGAGCCAAACCCG	1511
QY	201	AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr	220

SYNSELSTPRSDTLDLRQOLREITTEKTSKKNKDTLEKLOASGDALVDQBSQKETLENN
 SLPTTVTKTQAYGTVTNLQDTSVLQGNLQNMYSHNVLIMNLNLNLTLQVQQRN
 LITNLQNSQVQYATRKNDQNLQVFLQAKDQTDWLKVKVQSLQTLAANNALAL
 KANNITLEDNMSQLNSQVQYATRKNDQNLQVFLQAKDQTDWLKVKVQSLQTLA
 RQOLFDTDTVNIISNTSYTAHULRTLSQANEQLDKDLQDHKAENRTAIKFNQLEE
 RLDVSLRWQOOLMRSRLDTEAVLVSIMEEMKLVDKSHQGLIKNFTILOGPVCPGP
 RDRGVSQGPPTNGKQGEKEGPGPCPGERGPIGPAGPPCEGGKSGKSGQCPK
 GSRGSGKGPQGPSGDPPGPPGKXEGLEGPGQGPFGQLQGTGCEGVPVGPGRGLPG
 LQGVPMQPGKPPGPGPSGVALQNEPTPAPEDNSHCPHWNKFTDKCYFFSVVE
 KEIFEDAKLFCDEKDSHLVFINTRBEQQWIKQMVGRESHWIGLTPSERENKWLQD
 TSPDYKNKWAQOPDNNGHGGPGCEDCAGLIYAGOWNDFQCEDVNNFTICEKDRETVLSS
 AL"

polyA_signal 3020..3025
 /gene="SRCL"

BASE COUNT 911 a 744 c 764 g 639 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3 17e-58 Length: 3058
 Score: 1934.00 Matches: 341
 Percent Similarity: 99.71% Conservative: 0
 Best Local Similarity: 99.71% Mismatches: 1
 Query Match: 99.69% Indels: 0
 DB: 9 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AB038518 (1-3058)

Qy	1	MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle	20
Db	1409	ATGCAACAAGATTGTATGAGTTCGAGGTGACACTGAAGTAGGCAACTTATCATAGT	1468
Qy	21	MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle	40
Db	1469	ATGGAAGAATGAAGCTAGTAGACTCCACAAGCATGTCAGCTCATCAAGATTTTACAATA	1528
Qy	41	LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	60
Db	1529	CTACAAGGTCCACGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCACAGGACCCCT	1588
Qy	61	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro	80
Db	1589	GGCCCACTGGCAACAGGNCACAAAGAGAGAGGGGGAGCCTGGNACCACCTGGCCCT	1648
Qy	81	AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly	100
Db	1649	CGGGGTGAGAGAGCCCAATTGGACACAGCTGGTCCCCCGGAGAGCGCTGGCGGCAAGGA	1708
Qy	101	SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln	120
Db	1709	TCTAAAGCTCCCAAGGCCCCAAAGGCTCCCGTGTTCCTGGGAAGCCCGGCCCTCAG	1768
Qy	121	GlyProSerGlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyPro	140
Db	1769	GGCCCCAGTGGGACCCAGGCCCCCGGGCCCAACAGGAAAGAGGACTCCCCGGCCCT	1828
Qy	141	GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly	160
Db	1829	CAGGGCCCTCTGCTCTCCAGGGAATTCAGGGCACCTGTGGGGAGCCTGGGGTGCTTGA	1888
Qy	161	ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro	180
Db	1889	CCTCGGGACTGCCAGCTTGCCTTGGGTACCGAGCATGCCAGGCCCAAGGGCCCCCCC	1948
Qy	181	GlyProProGlyProSerGlyAlaValProLeuLeuAlaLeuGlnAsnGluProThrPro	200
Db	1949	GGCCCTCTGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAAGATGAGCCCAACCCG	2008
Qy	201	AlaProGluAspAsnGlyCysProProHisrPlysAsnPheThrAspLysCysTyr	220
Db	2009	GCACCGAGGACAATAGTGGCCCCCTCACATGGGAAGAACTTCACAGACAAATGCTACTAT	2068
Qy	221	PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer	240


```

Db 948 CATCTGTTTTCATAAACTAGAGGACACGCAATGGATAAAAACAGATGGTAGGG 1007
|||||
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluAtrGluAsnGluTrpLysTrpLeu 280
|||||
Db 1008 AGAGAGACCACTGATGGCTCCACAGACTCAGAGCGTGAATGGAAGTGGCTG 1067
|||||
Qy 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||
Db 1068 GATGGACATCCAGACTACAAAAATTGGAAGCTGGACAGCCGGAATAACTGGGTCAT 1127
|||||
Qy 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleThrAlaGlyGlnTrpAsnAspPhe 320
|||||
Db 1128 GGCCATGGCCAGGAGAGACTGCTGGTGGTGAATTTATGCTGGCAGTGGAAACGATTTC 1187
|||||
Qy 321 GlnCysGluAspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSer 340
|||||
Db 1188 CAATGTGAAGAGCTCAATACTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCT 1247
|||||
Qy 341 AlaLeu 342
|||||
Db 1248 GCATTA 1253

RESULT 7
AB078434
LOCUS
DEFINITION
AB078434
ACCESSION
AB078434.1
VERSION
AB078434.1
KEYWORDS
SOURCE
Mus musculus tissue_lib:Liver cdna to mRNA.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuo,A., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.
cdna cloning of mouse CL-P1 gene
Unpublished
2 (bases 1 to 2637)
Direct Submission
Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuo,A., Sakamoto,T. and Wakamiya,N.
Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College,
Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido
078-8510, Japan (E-mail: ohtanieasahikawa-med.ac.jp,
Tel:81-166-68-2393, Fax:81-166-68-2399)
FEATURES
Location/Qualifiers
1..2637
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_lib="Liver"
1..2637
/gene="CL-P1"
92..2320
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAC05523.1"
/db_xref="gi:21901969"
translation="MKDDFAEEVEVSFGYKRFQIQGTCTCKNNWALKFFSTVLVY
STNSELSTFRSDLDLROQLQEIETKTSKNDKLTAVESDLKLGPOACKAL
ILCLALLITVALIKYKVKEMNDVTGDMETSHQYDNKLTAVESDLKLGPOACKAL
SFLITVYNTKLTQANGYTNLQDTSVLQGNLQSNQYSQSVVIMNLLNLTQVQRN
LISNQSDVDTSLAIQRIKNDQNFQIARQKDDTLKQVDSQSLQTLAANSALA
KANNDLQSNLSSSTGQMDNTTISQANEQSLKDLQDLKHTENRTAVKFSQLEE
RFQVETDVIINIIISVTAHLRTLSNLDVRTCTDTLTRHTDDITLSUNNTLVNI
RLDSISLRMOODMRKLDTEVANSVYMEMLVDSKHGOLIKNFTILQGPGRGP
KDRGSGPGPGTKNGKQKBEKGPAGERTGIPVGPGRSGSKSGSQGPK
GSRGSPKPGQSGSGDPGPGPKDLPQPGPGFQGLQGTGVPFGPFGPGLPG
LPVFGPMGPKGPGPGSGMEPLALQNEPTPASEVNGCPHWNKFTDKCYFSLE
KEIFEDAKLFCEDKSSHLVFINREEQWIKKHTVVGRESHIGLITDSQESSEWKLGD

```

```

SPVDYKNWKAGQPDNMGSGHGPEDCAGLIYAGQWDFQCDEINNFICEKREAVPSS
IL"
BASE COUNT 829 a 617 c 635 g 556 t
ORIGIN
Alignment Scores:
Pred. No.: 9.37e-54 Length: 2637
Score: 1801.00 Matches: 313
Percent Similarity: 95.91% Conservative: 15
Best Local Similarity: 91.52% Mismatches: 14
Query Match: 92.84% Indels: 0
Db: 10 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AB078434 (1-2637)
Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||
Db 1292 ATGCAGCAGACATGATGAGGTCAAGATTAGACACTGAAGTGGCCAACTTATCAGTGGTT 1351
|||||
Qy 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhelThrIle 40
|||||
Db 1352 ATGGAAGAGATGAAACTGGTTGACTCCCAAGCACGGTCAGCTCATCAAGAACTTTACCAT 1411
|||||
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
|||||
Db 1412 CTACAAAGGTCTCTCTGGGCCAGAGGTCCAAAGGTGACAGAGGATCTCAGGAGCACCAT 1471
|||||
Qy 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
|||||
Db 1472 GGTCAACTGGCAACAAGGACAGAAAGAGAGAGAGGAGAGGAGGAGGAGGAGGAGGAGG 1531
|||||
Qy 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLys 100
|||||
Db 1532 GCGGCTGAGAGGGGACAACTTGACCAAGTCGCGCCCTCTCTGGAGAGCGTGGCAGCAAGCA 1591
|||||
Qy 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
|||||
Db 1592 TCCAAAGGCTCACAGGGTCCCAAGGATCTCGTGGGTCCCGGAGGAGGAGGAGGAGGAGG 1651
|||||
Qy 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
|||||
Db 1652 GGACCTAGTGGGGAGCCAGGACCAAGGTCACAGGTCACAGGCAAGGAGGAGGAGGAGGAGG 1711
|||||
Qy 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||
Db 1712 CAGGCGCTCTCTGGCTTCCAGGAGCTACAGGCGACTGTGGGTGAGCTGGAGTACCTGGA 1771
|||||
Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||
Db 1772 CCTCGGGGGTTGCCAGGCTTGCAGGGGTGCCAGGATGCCCTGGGCTTAAGGAGCACCAT 1831
|||||
Qy 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||
Db 1832 GGCCCTCCAGGCGCCCTCAGGAGCAATGGAGCCATTGGCTCTCGAGAATAAACAACCCCA 1891
|||||
Qy 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPhelThrAspLysCysTrpTrp 220
|||||
Db 1892 GCATCAGAGGTACACGGATGTCGCTCACTGGAAGAACTTCACAGATAAATGCTACTAT 1951
|||||
Qy 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
|||||
Db 1952 TTTTTCATTGGAAAAAGAAATTTTGAAGATGTAAGCTTTTCTGTGAAGACAAATCTTCC 2011
|||||
Qy 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
|||||
Db 2012 CATCTCGTTTTCATAAACTCAAGAGAAACAGCAATGGATGATAAAGCATACCGCTGGG 2071
|||||
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||
Db 2072 AGAGAAAGCCATTGATCGGCTTCACAGACTCAGAACAGCAAGCAATGGAGTGGCTA 2131
|||||
Qy 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||

```

```
Db 2132 GACGGTCACTCTTGTATTACAAAACTGGAAGCTGGACACCAAGATAACTAGGGGCACT 2191
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2192 GGCATGGCCAGGAGAACACTGTGCTGCTGATTACGCAGGACAGTGGATGACTTC 2251
QY 321 GlnCysGluAspValAsnAsnPhelLecCysGluLysAspArgGluThrValLeuSerSer 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2252 CAGTGTGATGAATCAATCAATTCATTCTTGAGAGAGAAAGGAGCGCATCATCATCC 2311
QY 341 AlaLeu 342
|||||
Db 2312 ATATTA 2317

RESULT 8
AB038519
LOCUS AB038519 3291 bp mRNA linear ROD 18-JAN-2002
DEFINITION Mus musculus src1 mRNA for scavenger receptor with C-type lectin,
complete cds.
ACCESSION AB038519
VERSION AB038519.1 GI:18146951
KEYWORDS Mus musculus cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Nakamura,K., Funakoshi,H., Tokunaga,F. and Nakamura,T.
Molecular cloning of a mouse scavenger receptor with C-type lectin
(SRCL)(1), a novel member of the scavenger receptor family
Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
21575692
2 (bases 1 to 3291)
Nakamura,K. and Nakamura,T.
Direct Submission
Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
Location/Qualifiers
1..3291
/organism="Mus musculus"
/db_xref="taxon:10090"
1..3291
/gene="src1"
77..2305
/gene="src1"
/codon_start=1
/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDDFAEEEVQSFYKRFGIHEGCTQCTKINNWLAKFSIVLLY
ILCALLTITVALGVKVKNDVSDGMETSHQYDNKLTAVEDSLKLGDOACKKAL
STNSELSTFRSDLLDLRQLOEITEKTSKNKDTLEKLAANGDSLVDROSQKETLQNN
SFLITVNTKLOANGYVYNNLOQDNLQMSQSVYINNINLNLTQVQRN
LISNQLSDSDTSLAIQRIKNDQNLQVFLQAKRDTLWKVKQSLQTLAANNLSALA
KANNDLDDMNSQLSSFTGMDNITTSIQANEQSLKLDQLHKDTEKNTAVKFSOLEE
RFOVETDIVANISNYSVAHLRLTSLNLDVMTCTDTLTRHTDILTSLNNTLVNI
RLDSILRMQDMRSLKLTDEVANLSVYVEEKLVDSKHGQLIKNFETILQGPGRGP
KDRGSQGPPTGNKGQKGKGPGRGPGRGRTIGPVGPGGERGSKGSGQPK
GSRSPGMPGPGPGPGPGPKDGLPGFPQGGTQVGEFVPGPRGLPG
LPVPGMPGPGPGPGPGAMEFLALQNEPTFASVNGCPPHWKNFTDRCYFSLF
KEILEDALKFCEDEKSHLVFINSRREQQWIKKTVGRSHWIGLTDLSQESSEWKLWG
SPVDYKNKNAQPDNWSGSHGPGEDCAGLIYAGQWINDFQCDIEINNFICEKREAVPSS
IL"
polyA_signal 3246..3251
/gene="src1"
BASE COUNT 1013 a 757 c 759 g 762 t
ORIGIN
Alignment Scores:
Pred. No.: 1.74e-53 Length: 3291
```

```
Score: 1795.00 Matches: 312
Percent Similarity: 95.61% Conservative: 15
Best Local Similarity: 91.23% Mismatches: 15
Query Match: 92.53% Indels: 0
Db: 10 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x AB038519 (1-3291)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1277 ATGCAGCAAGACATGATGAGGTCAAGATTAGACACTGAAGTGGCCAACTATTACAGTGGT 1336
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhThrIle 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1337 ATGGAAGAGATCAAACTGGTTGACTCCCAAGCAGCGGTCACTCATCAAGAACTTTACCAT 1396
QY 41 LeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1397 CTACAAGGTCCTCTCGGCCAGAGGTCCTCAAAAGGTGACAGAGGATCTCAGGAGCACCT 1456
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyPro 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1457 GGTCCAACCTGGCACAGGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1516
QY 81 AlaGlyGluArgGlyProTleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1517 GCGGGTGAGAGGGGCACAAATTTGGACCAGTCGCGCCCTCTCTGGAGAGCGTGGCAGCAA 1576
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1577 TCCAAGGCTCACAGGGTCCCAAGGATCTCGTGGGTCCCCAGGGGAAGCTTGGCCCTCAA 1636
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1637 GCACTAGTGGGAGCCAGGACCACCGTCCACCAGGCAAGATGAGTCTCCCTGGCCCT 1696
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1697 CAGGGCCCTCTCTGGGTTCACAGGAGCTTACAGGAGCTGTGGGTGAGCGTGTACCTGGA 1756
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1757 CCTCGGGGTTGCCAGGCTTGCCAGGGGTGCCAGGATGCCCTGGGCTTAAGGAGCACCT 1816
QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1817 GGCCTCCAGGCCCTTCAGGACCAATGGAGCCATTGGCTCTGCAGAAATCAACCAACCCA 1876
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPhThrAspLysCysTyrTyr 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1877 GCATCAGAGGTCAACGGATGTCCGCTCACTGGAAGAACTTCACAGATAAATGCTACTAT 1936
QY 221 PheSerValGluLysLutlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1937 TTTTCATTTGGAAGAAATAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTCC 1996
QY 241 HisLeuValPheIleAsnTrpArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1997 CATCTCGTTTTCATAAATCAAGAGAAAGAACAGCAATGATAAAAAAGCATACCGTGGG 2056
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2057 AGAGAAAGCCATTGGATCGGCTCACAGACTCAGAGACAGAGAAAGCAAGTGAAGTGGCTA 2116
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2117 GACGGTCACTGTTGATTACAAAACCTGGAAGCTGGCAACACAGATAACTGGGGCACT 2176
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2177 GGCCATGGCCAGGAGAGACTGTGCTGGCTTGATTTACGAGGAGACAGTGGGAATGACTTC 2236
QY 321 GlnCysGluAspValAsnAsnPhelLecCysGluLysAspArgGluThrValLeuSerSer 340
```

```
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 2237 CAGTGTGATGAATCAATAACTTCTTTGTGAGAGGAAAGGAGGAGTACCATCATCC 2296
Qy 341 AlaLeu 342
|||
Db 2297 ATATTA 2302

RESULT 9
AB052103 Homo sapiens SRC1 mRNA for scavenger receptor with C-type lectin
LOCUS AB052103.1 linear PRI 08-MAR-2001
DEFINITION type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS
SOURCE Homo sapiens cdna to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRC1), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
JOURNAL 21092718
MEDLINE 2 (bases 1 to 4330)
REFERENCE Nakamura,K. and Nakamura,T.
AUTHORS Direct Submission
TITLE Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
JOURNAL School of Medicine, Division of Biochemistry, Biomedical Research
Center; 2-2 Yamadaoka, suite, Osaka 565-0871, Japan
(E-mail:knakamur@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)
FEATURES
source Location/Qualifiers
1..4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..4330
/gene="SRC1"
3..1871
/gene="SRC1"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:13365553"
/translation="MKDDFAEEVEVSEFGKRFGIHSGTQCPKCHNNWALKFSILLY
ILCALITITVAILGYKVKRMNDVGMETSRYQYDDKLTAVESDLKGLDQTKKAI
STNSELSTFRSDIIDLRLQQLREITEKTSKNKDTLEKLAQSGDALVDQSOLTEKLN
SFLITTVNKTQLQAYGVYTNLQDTSVLQGNLQNMYSHNVVIMNLNLTQVQQRN
LITNLQSRVDTSQAIQRIKNDQNLQVLAQKDDTLKKEVQSLOTLAANNSALA
KANNDLTEDMNSQLNSETGOMENITTSQANEQNLKQDLHLKDAENRTAKFNQLEE
RFQLEFDIVNIIISNYTAHLRLTSLNNEVRITCTDTLTKHDDTLNUNTLANI
RLDSVMQODMLSRDLRDETVANLSVIMEEMKLVDKSHGQLIKNFITLQPPGPRG
RGDRSGQPGPTKNGKQEGEPGPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG
GSRGSPKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG
LPVPGPMGPGKPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPGPG
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN

Alignment Scores:
Pred. No.: 5,86e-49 Length: 4330
Score: 1664.00 Matches: 342
Percent Similarity: 39.13% Conservative: 0
Best Local Similarity: 39.13% Mismatches: 0
Query Match: 85.77% Indels: 532
DB: 9 Gaps: 1

US-09-763-712a-2_COPY_206_547 (1-342) x AB052103 (1-4330)

Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1203 ATGCAACAAGATTTTGTAGGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1262
```

QY 205 ----- 205
Db 2343 TTTTATGATTATGTGTATGTATGATGGACAGGGTCTCACTCTGTCAACCCAGGCTGGAAT 2402
QY 205 ----- 205
Db 2403 GACGAAGTCAAGTAGTGCAAAATAGCTCATTGCGACGTTTGACCTCCCTGGGCTTAGGTGA 2462
QY 205 ----- 205
Db 2463 TCTTCCCACTCAGACTCCTGCGTAGCTGGGACACAGGGGTATGCCACACACCTGGCTA 2522
QY 205 ----- 205
Db 2523 ATTTTATTTTCTTTTGTAGAAACAGGGTCTTGCTATGTTGCCAGGCTGGTCTTGA 2582
QY 205 ----- 205
Db 2583 ACTCTGGCTCAAGAACCCCTCGCTCAGCTCTTGGCAAGTGTGGGATTCCAGGTGG 2642
QY 205 ----- 205
Db 2643 GAGCCACTGCACCTGCTCTTGCACCTTGTTTATTTATAGGAGCCAGCCCTCAGAAACAA 2702
QY 205 ----- 205
Db 2703 ACACAATGCAAAACAATTTCTCTTCAAGGAAGCCTGCTGGCGCCCTGAGGCTCTTCCCTC 2762
QY 205 ----- 205
Db 2763 GTTTTACATCTCACTCGGAGATAGATGGGCTCCCTCGGTAGCTTCAATCAACATGGG 2822
QY 205 ----- 205
Db 2823 ATGGAACCTCTGCTATGATGACTACATTAATTTCTTTTGGAGAAAGGTGTAGATGT 2882
QY 205 ----- 205
Db 2883 TAGGTGCTCTTAGTCTTTTGATAATTAACCTTTGGCAAACTCGAATGTTGTTAG 2942
QY 205 ----- 205
Db 2943 CAACCAAGCTATTTTCTCTCTTTGACCAACCGCTATTCTACTTGATTTTGGTCGT 3002
QY 205 ----- 205
Db 3003 GAGGCCGGGTTATTTTGTGATCCGTGAGCCCATTTTGTGCTGGGTTCTCACATCC 3062
QY 205 ----- 205
Db 3063 AAGGTTCCTGGGTTCCGTTATGATCCAGCTGACGCTGAACATCTGCGCTCGCCCTGCT 3122
QY 205 ----- 205
Db 3123 TCTTGTGTTGCTTCTGTCATTTCTGTCATTGTCATTGGAAGCGGATCCAAACACAGAACA 3182
QY 205 ----- 205
Db 3183 AAGACCCAGTCCAAGATCCATCCAGATTATGTGTTTTCGGCTTGCTCTGCGAGTCCACT 3242
QY 205 ----- 205
Db 3243 TTGGGTCAFTGCTTGTGCGAGCAGCTCCGTGAACGTCAGCTGTGGGCTCTGTCAGGGAC 3302
QY 205 ----- 205
Db 3303 GCTCAGGCTGCTCCAGCGGGACACAGTTTGGGCTCTGAAACACAGCTGAAGAAGGC 3362
QY 206 -----GlyCyspro 208
Db 3363 AGAGTGTCTTCGTGATCTTTTAAATCAATGTTTTCACATTTTTCATCTCTAGGCTGCCG 3422
QY 209 ProHisTrpLysAsnPhetrAspLysCysTyrTyrPheSerValGluLysGluIlePhe 228

Db 3423 CCTCACTGGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTT 3482
QY 229 GluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArg 248
Db 3483 GAGGATGCAAACTTTTCTGTGAAGACAAGCTCTTCACATCTTGTGTTTTCATAAACACTAGA 3542
QY 249 GluGluGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeu 268
Db 3543 GAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCTC 3602
QY 269 ThrAspSerGluArgGluAsnGlnTrpLysTrpLeuAspGlyThrSerProAspTyrLys 288
Db 3603 ACAGACTTCAGACGCGTGAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAA 3662
QY 289 AsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCys 308
Db 3663 AATTGGAAGCTGGACAGCCGATTAACCTGGGTCTATGGCCATGGCCAGAGAAGACTGT 3722
QY 309 AlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAspPhe 328
Db 3723 GCTGGTGTGATTATGCTGGCAGTGGAACGATTTCCAATGTGAAGACGCTCAATAACTTC 3782
QY 329 IleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 342
Db 3783 ATTTGCAAAAAGACAGGAGAGACTACTGTCTATCTGCATTA 3824
RESULT 10
AC112416 130763 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-350K4, *** SEQUENCING IN PROGRESS
DEFINITION ***, 28 unordered pieces.
AC112416
AC112416.3 GI:21743403
VERSION HTG; HTGS, PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 130763)
AUTHORS Murzyn D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen C., Alb Brooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbaria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisai, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozado, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mashiney, E., McLeod, M.P., Meador, M., Mei, G., Metker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oraguanye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Slisson, I.,

Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tameris, A., Tameris, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmami, K., Vasquez, L., Vera V., Villalón, D., Vinson, R., Wang, Q., Williams, G., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, C., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72
73	74	75	76	77	78
79	80	81	82	83	84
85	86	87	88	89	90
91	92	93	94	95	96
97	98	99	100	101	102
103	104	105	106	107	108
109	110	111	112	113	114
115	116	117	118	119	120
121	122	123	124	125	126
127	128	129	130	131	132
133	134	135	136	137	138
139	140	141	142	143	144
145	146	147	148	149	150
151	152	153	154	155	156
157	158	159	160	161	162
163	164	165	166	167	168
169	170	171	172	173	174
175	176	177	178	179	180
181	182	183	184	185	186
187	188	189	190	191	192
193	194	195	196	197	198
199	200	201	202	203	204
205	206	207	208	209	210
211	212	213	214	215	216
217	218	219	220	221	222
223	224	225	226	227	228
229	230	231	232	233	234
235	236	237	238	239	240
241	242	243	244	245	246
247	248	249	250	251	252
253	254	255	256	257	258
259	260	261	262	263	264
265	266	267	268	269	270
271	272	273	274	275	276
277	278	279	280	281	282
283	284	285	286	287	288
289	290	291	292	293	294
295	296	297	298	299	300
301	302	303	304	305	306
307	308	309	310	311	312
313	314	315	316	317	318
319	320	321	322	323	324
325	326	327	328	329	330
331	332	333	334	335	336
337	338	339	340	341	342
343	344	345	346	347	348
349	350	351	352	353	354
355	356	357	358	359	360
361	362	363	364	365	366
367	368	369	370	371	372
373	374	375	376	377	378
379	380	381	382	383	384
385	386	387	388	389	390
391	392	393	394	395	396
397	398	399	400	401	402
403	404	405	406	407	4

REFERENCE	AUTHORS	TITLE	JOURNAL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRRG
Center clone name: CH230-350K4
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 113686 bases at least Q40
Consensus quality: 115729 bases at least Q30
Consensus quality: 117044 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 28 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1	1103: contig of 1103 bp in length	38	PheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGln	57
1104	1203: gap of unknown length		::	
1204	3068: contig of 1865 bp in length	13787	TTTCTTTATGTTTTAGTCTCTCTGGCCCCCAGAGGTCCAAAAGGTGACAGAGGATCTCAG	13846
3069	3168: gap of unknown length			
3169	4755: contig of 1587 bp in length	58	GlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro	77
4756	4855: gap of unknown length			
4756	4855: gap of unknown length			
4856	7046: contig of 2191 bp in length	13847	GNCCGCCCTGGTCCAACTGGCAACAAGGGACAGAAAGGAGAGAGAGAGCCCTGGTCCA	13906
7047	7146: gap of unknown length			
7147	8651: contig of 1505 bp in length	78	ProGlyProAlaGlyGluArgGlyProLleGlyProAlaGlyProGlyGlyGluArgGly	97
8652	8751: gap of unknown length			
8752	10001: contig of 1250 bp in length	13907	CCTGGCCCTGCGGGTGAGCGAGGCACAAATGGACCAGTAGGCCCTCTCTGGAGAGCGGTGGC	13966
10002	10101: gap of unknown length			
10102	11278: contig of 1177 bp in length	98	GlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro	117
11279	11378: gap of unknown length			
11379	13651: contig of 2273 bp in length	13967	AGCAAGGATCCAAAGGCTCACAGGGTCCCAAGGATCAGTGGGTCCCCAGGAAGGCT	14026
13652	13751: gap of unknown length			
13752	15942: contig of 2191 bp in length	118	GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeu	137
15943	16042: gap of unknown length			
16043	17619: contig of 1577 bp in length	14027	GGCCCTCAAGGGCCCTAGTGGAGACCCCTGGACCACCCAGGTCCACCAGGCAAGATGGACTC	14086
17620	17719: gap of unknown length			
17720	20301: contig of 2582 bp in length	138	ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly	157
20302	20401: gap of unknown length			
20402	23598: contig of 3197 bp in length	14087	CCTGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTACAGGGGACCTGTGGGTGAACCTGGA	14146
23599	23698: gap of unknown length			

QY 158 valProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys 177
Db 14147 GTACCTGGACCTCGGGGCTACAGGCTTACCAGGGGTGCCAGCATGCGTGGCCCAAG 14206
QY 178 GlyProGlyProProGlyProSerGlyAlaValProGlyLeuAlaLeuGlnAsnGlu 197
Db 14207 GGACCACTTGGCCCCCAGGTCTCCTCAGGAGCAATGGAGCCGTGCTCGCAATGAA 14266
QY 198 ProThrProAlaProGluAspAsn 205
Db 14267 CCAACCCAGCATCAGAGGTCAACGGTGAAGTCCAAATCATCTTTTCAGCCAGAAATAAG 14326
QY 205 ----- 205
Db 14327 GCAGTGTGTAGGCTTGGCTTAACACAGACCTCAGGAGTCCCTAAATTACACTTAATATCAG 14386
QY 205 ----- 205
Db 14387 GGAACCTATGTCTCTCTGGATGCGTGGTGTCTCACTCAGGACCCGTGTTCTCTGCT 14446
QY 205 ----- 205
Db 14447 TAAACATTATCAGGACTTACTCTATTAGATGAGCAGGCCCTCAGAAGATACCTATACC 14506
QY 205 ----- 205
Db 14507 CCAAGTACTATTTCAGAGTATCTCGACCGGCTGCTCCCTGACCATCTTCTCTGTCT 14566
QY 205 ----- 205
Db 14567 TCTTCTTTCTCTAGAAACAGACCATTCCATCAATTAACCTCATCAGTATGAGATGGG 14626
QY 205 ----- 205
Db 14627 TTTCTGTGTAGACCACTCTGGTTTGTGAAGAAGCTGCTGTGCTGTCTTTT 14686
QY 205 ----- 205
Db 14687 GACCATTAAAGTCTTCACAAAACCTCTGTTACTAGGAAACAAGCCATGCGGTCACTCA 14746
QY 205 ----- 205
Db 14747 TGCTGTGTCATGAACCTTTGGTCTGAAGTACACTGTTTACAGCTCCAGCTGATCCAAG 14806
QY 205 ----- 205
Db 14807 ACTGACAATGAGTATCAGGGCCACCCACCTTCTTCTTCTGTCATTATAAAGCCTAAC 14866
QY 205 ----- 205
Db 14867 CCAACCAATCAGACGCTTTCCACACTGTGTGTGTTCTACTATTGTGTACTATAAATCC 14926
QY 205 ----- 205
Db 14927 ACNACTGTGAACATCTGTCATGTTCTGTACAGGGACACTCCATTGGCATGCAAGCAGTG 14986
QY 205 ----- 205
Db 14987 TGAGGCTGTCACCTTGTGTATCTTAACATAGCTGAGGAAATCAGAGTGCCCTTTTTTTT 15046
QY 206 ----- GlyCysProProHisThrPlysAsnPheThr 215
Db 15047 AGATCAACATTTAAACACGTTTACATTCCTPAGGATGTCCGCCCTCCTCAGGAAGACTTCACA 15106
QY 216 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 235
Db 15107 GATAATGTTACTATTTTTCGGTGGGAAAGAAATTTTGAAGATGCTAAGCTTTTCTGT 15166
QY 236 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGln ----- 251
Db 15167 GAAGACAAATCTTCCCATCTCGTTTTCATATAAACTCAAGAGAGGAACAGTATGTACGCTG 15226
QY 252 ---GlnTrpIleLysLysGlnMetValGlyArg-----GluserHisrPleGly 267

Db 15227 CATACTTGTTGGAGAAAACATAATTTAGAGATAAGTTATATACTTGAATAATTCATGCGCT 15286
QY 268 LeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyr 287
Db 15287 TTGAGGAGTCGAAGAGATCAAAAATTCATGCTCTCATAAACAAGGAAGCTCAGTTTG 15346
QY 288 LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAsp 307
Db 15347 -----CATGGCTTTTATCCAGCATCGCA 15370
QY 308 CysAlaGlyLeu-----IleTyrAlaGlyGlnTrp 317
Db 15371 ATATGTGAATGTGTCATGCTTCTTAGGCCACAGATATAGTCCACCTGG 15421
RESULT 11
LOCUS AC114677 182029 bp DNA linear HTG 10-JUN-2002
DEFINITION Mus musculus clone RP24-213K19, WORKING DRAFT SEQUENCE, 10 ordered
AC114677 pieces.
VERSION AC114677.3 GI:21362159
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 182029)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-213K19
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 182029)
REFERENCE 1 (bases 1 to 182029)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 182029)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,

Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 10, 2002 this sequence version replaced gi:21328559.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24628
Center clone name: 213 K 19

```

----- Summary Statistics
Sequencing vector: plasmid; n/a: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179306 bases at least Q40
Consensus quality: 180773 bases at least Q30
Consensus quality: 181030 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 181129; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-fp
Quality coverage: 8.0 in Q20 bases; sum-of-contigs

```

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 951: contig of 951 bp in length
* 952 1051: gap of 100 bp
* 1052 3806: contig of 2755 bp in length
* 3807 3906: gap of 100 bp
* 3907 8709: contig of 4803 bp in length
* 8710 8809: gap of 100 bp
* 8810 17263: contig of 8454 bp in length
* 17264 17363: gap of 100 bp
* 17364 36938: contig of 19575 bp in length
* 36939 37038: gap of 100 bp
* 37039 53226: contig of 16188 bp in length
* 53227 53326: gap of 100 bp
* 53327 369572: contig of 16246 bp in length
* 69573 69672: gap of 100 bp
* 69673 89021: contig of 19349 bp in length
* 89022 89121: gap of 100 bp
* 89122 128310: contig of 39189 bp in length
* 128311 128410: gap of 100 bp
* 128411 182029: contig of 53619 bp in length.

```

FEATURES	source	Location/Qualifiers
		1. .182029
		/organism="Mus musculus"
		/db_xref="taxon:10090"
		/clone="RP24-213K19"
		/clone_lib="RPC1-24 Male Mouse BAC"
misc_feature		1. .951
		/note="assembly_fragment"
misc_feature		1052. .3806
		/note="assembly_fragment"
misc feature		3907. .8709

[illegible]

Db 45887 TGTTCTGCTTAGACATATACAGTCTCTACTCACTTTAGATGAGCCAGGCCTTCAGAAGA 46946
QY 203 ----- 203
Db 46947 CCCTGTACCCCAATACCTCTTCAAGATATCTACACTGGCTGCTCCCTGACCATCCT 47006
QY 203 ----- 203
Db 47007 CCTTCATCTTTCTCTTTCTCTCTTAGAAAAGACCATTTTATCAGTTAACTCCACCAGTA 47066
QY 203 ----- 203
Db 47067 TGCAGATAGGGTTCTGTGACAGAGTCCACATCTCTGGTTTTCCTTAAGAAGCTGTGTGG 47126
QY 203 ----- 203
Db 47127 TGCTTTCTGTCTCTTTGACAGAAAGCTTTTGCAAAACTATACGTAGTAGGAGCCAAAGC 47186
QY 203 ----- 203
Db 47187 GTGTGGTCCCTCATGCTGCTGCACAAAGCAGACTTTTGTCTCTGAAGTGTGCTTAFTTAC 47246
QY 203 ----- 203
Db 47247 AACCTCCAGCGATCCAAAGCCGACAGTGTATCAGAACCCACCCACTTCTTTTCTCTGT 47306
QY 203 ----- 203
Db 47307 CGTCATAAGCCTAGCCCAACCCAGTCCACCTTCTTCTGCTGTGGTGTTCACGTGCTG 47366
QY 203 ----- 203
Db 47367 TGTGCTGTACAGTCCACATCTGAGAACATCTGCCGGGTTCTACACAGGACACCCATTG 47426
QY 203 ----- 203
Db 47427 GCATGCAAGTGTACAAGCTGTGACTTTGATATCTTAACACACAGCTGAGGAATGCAGG 47486
QY 204 ----- AspAsnGlyCysProPro 209
Db 47487 GTTCCTTCTGTGTTTTTAATCAACAGTAAACTTTTACATTCCTA-CGATGTCGGCT 47545
QY 210 HisTrpLysAsnPhetThrAspLysCysTyrPheSerValGluLysGluLeuPheGlu 229
Db 47546 CACTGGAAGAACTTCACAGATAAACTACTATTTTTCATTGGAAGAAATTTTGA 47605
QY 230 AspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGlu 249
Db 47606 GATGCTAAGCTTTCTGTGAAGACAAATCTCCCATCTCGTTTTCATAACTCAAGAGAA 47665
QY 250 GluGlnGln-----TrpIleLysLysGlnMetValGlyArgGlu 262
Db 47666 GAACAGGT-ATGATATGTGACGTGCACACTTGTGTGGAGAAACTATATTAGAGATA- 47723
QY 263 SerHisTrpIleClyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGly 282
Db 47724 AGTTATACTTGGAAATGTGATGCTTTTGAGAGAGTCAAAAGATCAA----- 47771
QY 283 ThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHis 302
Db 47772 -----AATTCAATGGTCTCATAAACAA 47792
QY 303 GlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrp---AsnAspPheGln 321
Db 47793 AGGAAGCTGCAGTTTGCATGGCTTTTCCAGGATGAGAGTATGATGATGAATCAAA 47852
RESULT 12
AC102618
LOCUS AC102618 193208 bp DNA linear HTG 21-AUG-2002
DEFINITION Mus musculus clone RP23-426G16, WORKING DRAFT SEQUENCE, 11
unordered pieces.
AC102618
AC102618

VERSION
KEYWORDS
SOURCE
ORGANISM

AC102618.2 GI:22381604

HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

house mouse.

Mus musculus

REFERENCE

AUTHORS

1 (bases 1 to 193208)
Birren,B., Nusbaum,C. and Lander,E.

TITLE

Mus musculus, clone RP23-426G16

REFERENCE

AUTHORS

2 (bases 1 to 193208)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangeo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 193208)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tefsfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 21, 2002 this sequence version replaced gi:17061704.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9116

Center clone name: 426_G16

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 190144 bases at least Q40
Consensus quality: 191397 bases at least Q30
Consensus quality: 191862 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 192208; sum-of-contigs
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 9.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

- * 1 1213: contig of 1213 bp in length
- * 1214 1313: gap of 100 bp
- * 1314 3101: contig of 1788 bp in length
- * 3102 3201: gap of 100 bp
- * 3202 6759: contig of 3558 bp in length
- * 6760 6859: gap of 100 bp
- * 6860 9034: contig of 2175 bp in length
- * 9035 9134: gap of 100 bp
- * 9135 74024: contig of 64890 bp in length
- * 74025 74124: gap of 100 bp
- * 74125 88784: contig of 14660 bp in length
- * 88785 88884: gap of 100 bp
- * 88885 107450: contig of 18566 bp in length
- * 107451 107550: gap of 100 bp
- * 107551 127842: contig of 20292 bp in length
- * 127843 127942: gap of 100 bp
- * 127943 150179: contig of 2237 bp in length
- * 150180 150279: gap of 100 bp
- * 150280 188446: contig of 38167 bp in length
- * 188447 188546: gap of 100 bp
- * 188547 193208: contig of 4662 bp in length.

FEATURES

source
1..193208
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RP23-426G16"
/clone_lib="RPC1-23 Female Mouse BAC"
1..1213
/note="assembly_fragment"
1314..3101
/note="assembly_fragment"
3202..6759
/note="assembly_fragment"
6860..9034
/note="assembly_fragment"
9135..74024
/note="assembly_fragment"
74125..88784
/note="assembly_fragment"
88885..107450
/note="assembly_fragment"
107551..127842
/note="assembly_fragment"
127943..150179
/note="assembly_fragment"
150280..188446
/note="assembly_fragment"
188547..193208
/note="assembly_fragment
clone_end:T7
vector_side:right"

BASE COUNT 54660 a 40174 c 40217 g 57157 t 1000 others

ORIGIN
Alignment Scores:
Pred. No.:

1.18e-24 Length: 193208

Score:	981.00	Matches:	214
Percent Similarity:	41.25%	Conservative:	17
Best Local Similarity:	38.21%	Mismatches:	35
Query Match:	50.57%	Indels:	297
DB:	2	Gaps:	5
US-09-763-712A-2_COPY_206_547 (1-342) x AC102618 (1-193208)			
Qy	38	PheThrIleLeuGlnGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGln	57
Db	97943	TTTTATGTTTAA---GGTCCCTCTGGCCCGCCAGAGTCCAAAAGGTGACAGAGATCTCAG	97999
Qy	58	GlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro	77
Db	98000	GGACCACTGGTCCCAACTGGCAACAAGGACAGAAAGGAGAGAGAGAGAGCTGGTCCA	98059
Qy	78	ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGly	97
Db	98060	CCTGGCCCTTGGCGGTGAGAGGGGCACAATTGGACAGATCGGCCCTCTCGAGAGCGTGC	98119
Qy	98	GlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro	117
Db	98120	AGCAAGGATCCAAAGGCTCACAGGTCCTCCAAAGGATCTCGTGGTCCCGGAGGACCT	98179
Qy	118	GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeu	137
Db	98180	GGCCCTCAAGGACCTAGTGGGGACCCAGGACCCAGGTCACAGGAGGATGGACTC	98239
Qy	138	ProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly	157
Db	98240	CCTGGCCCTTCAGGGCCCTCTGGCTTCCAGGACATACAGGGCAGCTGTGGGTGAGCCTGA	98299
Qy	158	ValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys	177
Db	98300	GTACCTGGACCTCGGGGTTGCCAGGCTTGCAGGGTGCCAGGATGCTGGGCTTAG	98359
Qy	178	GlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGlu	197
Db	98360	GGACCACTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCATTTGGCTCTGCAGATGAA	98419
Qy	198	ProThrProAlaProGlu-----	203
Db	98420	CCACCCCGACGATCAGAGGTCAACGGTAAGTCCAATCGCTCTTCAGCTAGAATTGTG	98479
Qy	203	-----	203
Db	98480	GAGGAATAAGGCAGTGTGTAGGCTTGGCTTAAACCAACCCCGAGGAGTCCCTAATTACAC	98539
Qy	203	-----	203
Db	98540	AAACACAGCAGGAGCCCTGTGTATCCCTCTGGATGCTTCACTGCTGTCTACTCAGGACCCCTG	98599
Qy	203	-----	203
Db	98600	TGTTCCCTGCTTACATATTATCAGGCTCTTACTCATTTAGATGAGCCAGCCCTTCAGAAGA	98659
Qy	203	-----	203
Db	98660	CCCTGTACCCCAATACTTCTTCAAGATATTCTACATTTGCTGGCTGCTCCCTGACCATCT	98719
Qy	203	-----	203
Db	98720	CCTTCATCTTCTTCTTCTTCTTCTTAGAACAGACCATTTTATCAGTTAACTCCACCAGTA	98779
Qy	203	-----	203
Db	98780	TGCAGATAGGTTTCTGTGACAGAGTCCACATTTCTGGTTTTTCTTAAGAAGCTGTTGTGG	98839
Qy	203	-----	203
Db	98840	TGCCTTCTGTCTCTTGTGACAGAAAGTCTTTTGCATAAACTATACTAGTTAGGAGCCACAC	98899
Qy	203	-----	203

```
Db 98900 GTGTGGTCCCTCATGCTGGTCCCAAGCCAGCACTTTGGTTCTGAAGTGTGCTATTATAC 98959
QY 203 ----- 203
Db 98960 AACCTCCAGCCGATCCAGCCGACAGTGTATCAGAACCCACCCCACTTCTTTTTCGTGT 99019
QY 203 ----- 203
Db 99020 CGTCATAAGCCCTAGCCCAACCCAGTCCCAACCTTCTTCTTCACTGTGTGGTTCACCTGCTG 99079
QY 203 ----- 203
Db 99080 TGTGCTGTACATCCACATCTGAGAACATCTGCCGGGTTCTACACAGGACACCCATTG 99139
QY 203 ----- 203
Db 99140 GCATCAAGTGTACAGGCTGTGACTTTGATATCTTAACACAGCTGAGGAATGCGAGG 99199
QY 204 ----- AspAsnGlyCysProPro 209
          |||||
Db 99200 GTTCTCTTGTGTTTTTAAATCAACAGTTAAACTTTTACATTCCTA-GGATGTCGCGCT 99258
QY 210 HistTrpLysAsnPhetAspLysCysTyrTyrrPheSerValGluLysGluIlePheGlu 229
          |||||
Db 99259 CACTGGAAGAACTTCACAGATAAATGCTACTATTTTTCATTGCAAAAAGAAATTTTGAA 99318
QY 230 AspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGlu 249
          |||||
Db 99319 GATGCTAAGCTTTCTGTGAACACAAATCTCCCATCTCGTTTTCATAAATCAGAGAA 99378
QY 250 GluGlnGln-----TrpIleLysLysGlnMetValGlyArgGlu 262
          |||||
Db 99379 GAACAGGT-ATGATATGTGAGTGCACACTTGGTTGGAGAAACTATAATTTAGAGATA- 99436
QY 263 SerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGly 282
          |||||
Db 99437 AGTTATACTTGGAAATGATGCCTTTTGAGAGAGTCAAGAGATCAAA----- 99484
QY 283 ThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHis 302
          |||||
Db 99485 -----AATCATGGTCTCTATAACAA 99505
QY 303 GlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrp---AsnAspPheGln 321
          |||||
Db 99506 AGGAAGGCTGCAGTTTGCATGGGCTTTTCCAGGTATGAGAGTATGAATGAATTCAAA 99565
RESULT 13
AC016128/c 169088 bp DNA linear HTG 09-SEP-2000
LOCUS Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 8 unordered pieces.
ACCESSION AC016128
VERSION AC016128.4 GI:10046526
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169088)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
          Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B.,
          Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
          Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doneian,L., Doyle,M.,
          Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
          Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
          Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
          McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
```

```
TITLE
JOURNAL
COMMENT
...
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 1999
Center clone name: 324_G_2
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165522 bases at least Q40
Consensus quality: 167248 bases at least Q30
Consensus quality: 167996 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 168388; sum-of-contigs
Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17943: contig of 17943 bp in length
* 17944 18043: gap of 100 bp
* 18044 22424: contig of 4381 bp in length
* 22425 22524: gap of 100 bp
* 22525 38094: contig of 15570 bp in length
* 38095 38194: gap of 100 bp
* 38195 49220: contig of 11026 bp in length
* 49221 49320: gap of 100 bp
* 49321 66394: contig of 17074 bp in length
* 66395 66494: gap of 100 bp
* 66495 91692: contig of 25198 bp in length
* 91693 91792: gap of 100 bp
* 91793 120869: contig of 29077 bp in length
* 120870 120969: gap of 100 bp
* 120970 169088: contig of 48119 bp in length.
FEATURES
          source
          1..169088
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="18"
            /map="18"
            /clone="RP11-324G2"
            /clone_lib="RP11-11 Human Male BAC"
          1..17943
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left
            18044..22424
              /note="assembly_fragment"
              22525..38094
                /note="assembly_fragment"
                38195..49220
                  /note="assembly_fragment"
```

```
misc_feature 49321..66394
/note="assembly_fragment"
misc_feature 66495..91692
/note="assembly_fragment"
misc_feature 91793..120869
/note="assembly_fragment"
misc_feature 120970..169088
/note="assembly_fragment"
clone_end:T7
vector_size:right"
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN
Alignment Scores:
Pred. No.: 8.52e-24 Length: 169088
Score: 954.50 Matches: 191
Percent Similarity: 58.40% Conservative: 21
Best Local Similarity: 52.62% Mismatches: 53
Query Match: 49.20% Indels: 98
DB: 2 Gaps: 8
US-09-763-712A-2_COPY_206_547 (1-342) x AC016128 (1-169088)
Qy 16 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGly-----Gln 33
||||| : : : : : |||||
Db 153903 AATTGGAGATTATTATAAGAGAAAAATATTCAACTATCATTCACAGTGGATTTTCAT 153844
Qy 34 LeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAsp 53
: : : : : |||||
Db 153843 GTTGACACTTTTCTTTTATGCTTTAGTCCACCGGCCCCAGGGGTCGAAGAGTGAC 153784
Qy 54 ArgGlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGly 73
||||| : : : : : |||||
Db 153783 AGAGGATCCAGGGACCCCTGGCCCAACTGCAACAGGACAGAAAGAGAGAGGGG 153724
Qy 74 GluProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProPro 93
||||| : : : : : |||||
Db 153723 GAGCCTGGACCACTTGGCCCTCGGGGTGAGAGAGCCCAATTGGACCACTGGTCCGCC 153664
Qy 94 GlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySer 113
||||| : : : : : |||||
Db 153663 GGAGAGCTGGCGGCAAGGATCTAAGGCTCCCAAGGCCCAAGGCTCCCGTGGTTCC 153604
Qy 114 ProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProProGly 133
||||| : : : : : |||||
Db 153603 COTGGAAAGCCCGGCCCTCAGGCTCCAGTGGGGACCCAGGCCCCCGGGCCACCAAGC 153544
Qy 134 LysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrVal 153
||||| : : : : : |||||
Db 153543 AAAGAGGAGTCCCGGCCCTCAGGCGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGT 153484
Qy 154 GlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMet 173
||||| : : : : : |||||
Db 153483 GGGGAGCTGGGTGCTTGGACCTCGGGACTTGGGAGTGGCGGGGTACAGGCAATG 153424
Qy 174 ProGlyProLysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAla 193
||||| : : : : : |||||
Db 153423 CCAGGCCCAAGGGCCCCCGGCCCTCTTGGCCCATCAGGAGCGGTGGTGGCCCTGGCC 153364
Qy 194 LeuGlnAsnGluProThrProAlaProGluAspAsnGly----- 206
||||| : : : : : |||||
Db 153363 CTGCAAGATGAGCCCAACCCAGCACCGGAGGACAAATGTAAGTCCAAAGCCCTCCCTCCAG 153304
Qy 207 -----CysPro-ProHi 210
||||| : : : : : |||||
Db 153303 CCAGGGGACAGGGCAGTGCTGTGCTGTGAGACTGAGCCACCCCACTGTGTGCCCA 153244
Qy 210 sTriPysAsnPheThrAspLysCysTyTyTyPheSerValGluLysGluIlePheGluAs 230
||||| : : : : : |||||
Db 153243 C-----CCATGTTATTT----- 153231
Qy 230 pAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGl 250
||||| : : : : : |||||
```

Qy	16	AsnLeuSerValIleMetGluGluWetLysLeuValAspSerLysHisGly-----Gln	33
Db	45024	AATTTGGAGATTATTAAGAAGAAAAATTATCATTAACATATCATTCACAGTGGATTTCAT	44965
Qy	34	LeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAsp	53
Db	44964	GTTCACACCTTTTCTTTTATGCTTTAGTTCACCGGGCCCGAGGGTCCCAAGAGGTGAC	44905
Qy	54	ArgGlySerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGly	73
Db	44904	AGAGATCCCAAGGACCCCTGGCCCACTGGCAACAGGACAGAAAGGAGAGAGAGGGG	44845
Qy	74	GluProGlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProPro	93
Db	44844	GAGCTGACACACTGGCCCTCGGGTGGAGAGAGGCCCAATTGGACCACTGGTCCCCCC	44785
Qy	94	GlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySer	113
Db	44784	GGAGAGCGTGGCGCAAGGATCTAAAGGCTCCAGAGGCCCAAGGCTCCCGTGGTTC	44725
Qy	114	ProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGly	133
Db	44724	CCTGGGAAGCCCGCCCTCAGGGCTCCAGTGGGACCCAGGCCCCCGGGGCCACAGGC	44665
Qy	134	LysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrVal	153
Db	44664	AAAGAGGGACTCCCGCGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTT	44605
Qy	154	GlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMet	173
Db	44604	GGGAGAGCTGGGTGCTTGACCTCGGGACTCGGGAGCTGCAGGCTTGCTGGGGTACCAAGCATG	44545
Qy	174	ProGlyProLysGlyProProGlyProProGlyProSerGlyAlaValProLeuAla	193
Db	44544	CCAGGCCCAAGGCCCCCGCGCCCTCTGGCCCATCAGGAGCGGTGCTGCCCTGGCC	44485
Qy	194	LeuGlnAsnGluProThrProAlaProGluAspAsnGly-----	206
Db	44484	CTGCAGATGAGCCAAACCCAGCACCGAGGACAAATGGTAAGTCCAGGCCCTCCCTCCAG	44425
Qy	207	-----Cyspro-ProH1	210
Db	44424	CCAGGGGACAGGCAGTGCATGTCTTGAGACTGAGCCAAACCCCGAGTGTGGCCCA	44365
Qy	210	strLysAsnPheThrAspLysCysTyrtYrPheSerValGluLysGluIlePheGluAs	230
Db	44364	C-----CCATGCTATTTT-----	44352
Qy	230	pAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluG	250
Db	44351	-TGTTTTTTGTTTTGTTTTGAGACGGGTCTCACTCTGTT-----	44313
Qy	250	uGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAs	270
Db	44313	-----	44313
Qy	270	pSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTr	290
Db	44312	-GCCAGGCTGGAGTCAGTGGCTGTATCTCGGCTCACTGCAACTCCACATCCCGGGTT	44254
Qy	290	pLys-----AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyG	306
Db	44253	CAAGTGATTCCTCGCTAGCCCTCCAGGTAGCTGGGACTACAAGCTGTGCCCCACACA	44194
Qy	306	uAspCysAlaGlyLeuIleTyrAlaGlyGln-----	318
Db	44193	GCCAGCTAAATTTTGTATTTTGTAGAGAGGGGGTTCACCGTGTGGCCAGGCTGCTC	44134
Qy	318	n-----AspPheG	321
Db	44133	TCAAACCTCGGCCCTCAAGTGATCCCACTACCTTAGTCTCCCAAAGTGTGGGATTACAG	44074
Qy	321	nCysGlu	323

* 2385 3056: contig of 672 bp in length
* 3057 3156: gap of 100 bp
* 3157 3851: contig of 695 bp in length
* 3852 3951: gap of 100 bp
* 3952 4627: contig of 676 bp in length
* 4628 4727: gap of 100 bp
* 4728 5409: contig of 682 bp in length
* 5410 5509: gap of 100 bp
* 5510 6183: contig of 674 bp in length
* 6184 6283: gap of 100 bp
* 6284 6984: contig of 701 bp in length
* 6985 7084: gap of 100 bp
* 7085 7768: contig of 684 bp in length
* 7769 7868: gap of 100 bp
* 7869 8550: contig of 682 bp in length
* 8551 8650: gap of 100 bp
* 8651 9352: contig of 702 bp in length
* 9353 9452: gap of 100 bp
* 9453 10140: contig of 688 bp in length
* 10141 10240: gap of 100 bp
* 10241 10922: contig of 682 bp in length
* 10923 11022: gap of 100 bp
* 11023 11708: contig of 687 bp in length
* 11710 11809: gap of 100 bp
* 11810 12511: contig of 702 bp in length
* 12512 12611: gap of 100 bp
* 12612 13304: contig of 693 bp in length
* 13305 13404: gap of 100 bp
* 13405 14095: contig of 691 bp in length
* 14096 14195: gap of 100 bp
* 14196 14881: contig of 686 bp in length
* 14882 14981: gap of 100 bp
* 14982 15674: contig of 693 bp in length
* 15675 15774: gap of 100 bp
* 15775 16460: contig of 686 bp in length
* 16461 16560: gap of 100 bp
* 16561 17250: contig of 690 bp in length
* 17251 17350: gap of 100 bp
* 17351 18040: contig of 690 bp in length
* 18041 18140: gap of 100 bp
* 18141 18836: contig of 696 bp in length
* 18837 18936: gap of 100 bp
* 18937 19621: contig of 685 bp in length
* 19622 19721: gap of 100 bp
* 19722 20425: contig of 704 bp in length
* 20426 20525: gap of 100 bp
* 20526 21224: contig of 699 bp in length
* 21225 21324: gap of 100 bp
* 21325 22017: contig of 693 bp in length
* 22018 22117: gap of 100 bp
* 22118 22815: contig of 698 bp in length
* 22816 22915: gap of 100 bp
* 22916 23609: contig of 694 bp in length
* 23610 23709: gap of 100 bp
* 23710 24402: contig of 693 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25981: contig of 690 bp in length
* 25982 26081: gap of 100 bp
* 26082 26774: contig of 693 bp in length
* 26775 26874: gap of 100 bp
* 26875 27563: contig of 689 bp in length
* 27564 27663: gap of 100 bp
* 27664 28355: contig of 692 bp in length
* 28356 28455: gap of 100 bp
* 28456 29145: contig of 690 bp in length
* 29146 29245: gap of 100 bp
* 29246 29943: contig of 698 bp in length
* 29944 30043: gap of 100 bp
* 30044 30745: contig of 702 bp in length
* 30746 30845: gap of 100 bp
* 30846 31535: contig of 690 bp in length

* 31536 31635: gap of 100 bp
* 31636 32329: contig of 694 bp in length
* 32330 32429: gap of 100 bp
* 32430 33114: contig of 685 bp in length
* 33115 33214: gap of 100 bp
* 33215 33898: contig of 684 bp in length
* 33899 33998: gap of 100 bp
* 33999 34694: contig of 696 bp in length
* 34695 34794: gap of 100 bp
* 34795 35486: contig of 692 bp in length
* 35487 35586: gap of 100 bp
* 35587 36292: contig of 706 bp in length
* 36293 36392: gap of 100 bp
* 36393 37095: contig of 703 bp in length
* 37096 37195: gap of 100 bp
* 37196 37872: contig of 677 bp in length
* 37873 37972: gap of 100 bp
* 37973 38663: contig of 691 bp in length
* 38664 38763: gap of 100 bp
* 38764 39449: contig of 686 bp in length
* 39450 39549: gap of 100 bp
* 39550 40236: contig of 687 bp in length
* 40237 40336: gap of 100 bp
* 40337 41025: contig of 689 bp in length
* 41026 41125: gap of 100 bp
* 41126 41818: contig of 693 bp in length
* 41819 41918: gap of 100 bp
* 41919 42608: contig of 690 bp in length
* 42609 42708: gap of 100 bp
* 42709 43399: contig of 691 bp in length
* 43400 43499: gap of 100 bp
* 43500 44169: contig of 670 bp in length
* 44170 44269: gap of 100 bp
* 44270 44966: contig of 697 bp in length
* 44967 45066: gap of 100 bp
* 45067 45750: contig of 684 bp in length
* 45751 45850: gap of 100 bp
* 45851 46545: contig of 695 bp in length
* 46546 46645: gap of 100 bp
* 46646 47336: contig of 691 bp in length
* 47337 47436: gap of 100 bp
* 47437 48123: contig of 687 bp in length
* 48124 48223: gap of 100 bp
* 48224 48899: contig of 676 bp in length
* 48900 48999: gap of 100 bp
* 49000 49672: contig of 673 bp in length
* 49673 49772: gap of 100 bp
* 49773 50465: contig of 693 bp in length
* 50466 50565: gap of 100 bp
* 50566 51254: contig of 689 bp in length
* 51255 51354: gap of 100 bp
* 51355 52062: contig of 708 bp in length
* 52063 52162: gap of 100 bp
* 52163 52847: contig of 685 bp in length
* 52848 52947: gap of 100 bp
* 52948 53646: contig of 699 bp in length
* 53647 53746: gap of 100 bp
* 53747 54409: contig of 663 bp in length
* 54410 54509: gap of 100 bp

Alignment Scores:

Pred. No.: 2.09e-23 Length: 71044
Score: 935.50 Matches: 165
Percent Similarity: 92.43% Conservative: 6
Best Local Similarity: 89.19% Mismatches: 11
Query Match: 48.22% Indels: 3
DB: 2 Gaps: 1

US-09-763-712a-2_COPY_206_547 (1-342) x AC024368 (1-71044)

Qy 25 LysLeuValAspSerLysHisGlyGlnLeuIle-----LysAsnPhetrIleLeu 41
|||||: : : ||| ||| : : :
Db 43544 AAAATTATTCACTAACTATCATTCACAGTGGATTTCATGTTGACGGGGGGCTTTTAGCA 43603

QY 42 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 61
 Db 43604 AAAAGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGGATCCACAGGACCCCTGGC 43663
 QY 62 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 81
 Db 43664 CCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCTGGC 43723
 QY 82 GlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySer 101
 Db 43724 GGTGAGAGAGCCCGGGGACCACTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 43783
 QY 102 LysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGly 121
 Db 43784 AAAGGCTCCACAGGGCCCAAGAGCTCCCGTGTTCCTCTGGGAAGCGCCCTCAGGCGC 43843
 QY 122 ProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 141
 Db 43844 TCCAGTGGGGACCCAGGGCCCGGGCCCAAGAGGAGGACTCCCGGGCCCTCAG 43903
 QY 142 GlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyPro 161
 Db 43904 GSCCTCTCTGGCTTCCAGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGGCTGGACCT 43963
 QY 162 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 181
 Db 43964 CGGGGACTGCCAGGCTTGGCTGGGTATCCAGGCATGCCAGGCCCAAGGGCCCGCCCGGC 44023
 QY 182 ProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAla 201
 Db 44024 CCTCTGGCCCAATCAGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAACCCAGCA 44083
 QY 202 ProGluAspAsnGly 206
 Db 44084 CCGAGGACAATGGT 44098

Search completed: March 21, 2003, 11:21:28
 Job time : 2501.11 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run On: March 21, 2003, 02:20:06 ; Search time 179.385 Seconds
(without alignments)
4293.470 Million cell updates/sec

Title: US-09-763-712a-2_COPY_206_547

Perfect score: 1940

Sequence: 1 MQQDLMSRLDTEVANLSVI.....EDVNNFICKDRETVLSSAL 342

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n_model -DEV=xlp
-Q=cqn2_1/USPRO-spool/US09763712/runat_14032003_100949_18091/app_query.fasta_1.1877
-DB=N_Geneseq_101002 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.ecgn_1.1.626 @runat_14032003_100949_18091 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1940	100.0	2024	21	AAA07697 Human collectin en
2	1940	100.0	2262	24	ABA97932 Human scavenger re
3	1940	100.0	2318	22	AAI58842 Human polynucleoti
4	1940	100.0	2628	22	AAH43036 Nucleotide sequenc
5	1934	99.7	2005	24	ABL95574 Human anglogenesis
6	1934	99.7	2005	24	ABL88085 Human PRO7223 CDNA
7	1934	99.7	2181	22	AAI60628 Human polynucleoti
8	1934	99.7	2641	22	AAI60628 Human EXMAD-14 cod
9	1934	99.7	2929	24	ABQ92072 Human polynucleoti
10	1934	99.7	2930	20	AAV55746 Human secreted pro
11	1801	92.8	2637	22	AAH43037 Nucleotide sequenc
12	1715.5	88.4	1521	23	AAH71133 DNA encoding novel
13	1158	59.7	2256	22	AAH43054 Nucleotide sequenc
14	540	27.8	3685	20	AAH27858 Human CSR1 protein
15	540	27.8	3810	20	AAH27856 Human CSR1 protein
16	474.5	24.5	1410	24	ABK84622 Human CDNA differe
17	474.5	24.5	1410	24	ABL69664 Lung cancer relate
18	471	24.3	5676	15	AAQ64556 Human collagen (Ty
19	471	24.3	8284	22	AAAL26526 Human breast cance
20	471	24.3	8284	22	AAAL26553 Human breast cance
21	471	24.3	8284	22	AAAL26600 Human breast cance
22	462	23.8	4428	22	AAAD06574 Bovine alphas(III)
23	462	23.8	4428	22	AAAD06575 Bovine alphas(III)
24	459.5	23.7	756	14	AAQ43034 Collagen-like poly
25	459.5	23.7	756	17	AAAT16768 Collagen-like poly
26	458.5	23.6	4821	23	AAAS86866 DNA encoding novel
27	454	23.4	3171	21	AAAL2503 DNA encoding a hu
28	454	23.4	3171	21	AAZ99843 DNA encoding human
29	453	23.4	6158	24	ABL62095 Colon adenocarcino
30	453	23.4	6158	24	ABL65452 Lung cancer relate
31	453	23.4	6158	24	ABK35486 Human endometrial
32	452.5	23.3	3170	21	AAAL2493 CDNA encoding huma
33	452.5	23.3	3171	21	AAAL2502 CDNA encoding a hu
34	452.5	23.3	3181	19	AAV59358 Nucleotide sequenc
35	452.5	23.3	3349	17	AAAT1518 Collagen Al/decori
36	452.5	23.3	3349	21	AAAL2500 CDNA encoding a ch
37	452.5	23.3	3531	21	AAAL2497 CDNA encoding a ch
38	452.5	23.3	3535	17	AAAT1515 Collagen Al/TGF-be
39	452.5	23.3	3541	17	AAAT1516 CDNA encoding a ch
40	452.5	23.3	3541	21	AAAL2498 Collagen Al/TGF-be
41	452.5	23.3	4192	17	AAAT1517 Collagen Al/decori
42	452.5	23.3	4409	19	AAV60814 Human recombinant
43	452.5	23.3	4770	22	AAAS22441 Human CDNA encodin
44	452.5	23.3	6728	22	AAF90491 Human pro-alpha-1
45	452.5	23.3	6728	24	ABN97451 Gene #3949 used to

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX AAA07697;
AC AAA07697;
XX 20-JUN-2000 (first entry)
DT 20-JUN-2000 (first entry)
XX Human collectin encoding DNA.
DE Human collectin encoding DNA.
XX Collectin; human; antibacterial; antiviral; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 55..1698
FT /*tag= a
FT /product= "collectin"

AC ABA97932;
XX
DT 25-APR-2002 (first entry)
XX
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
KW Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
KW rheumatoid arthritis; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 33..2262
FT /tag= a
FT /product= "scavenger receptor-like protein"
XX
PN JP2001340089-A.
XX
PD 11-DEC-2001.
XX
PF 08-DEC-2000; 2000JP-0375066.
XX
PR 27-MAR-2000; 2000JP-0090772.
XX
PA (SHIO) SHIONOGI & CO LTD.
XX
DR WPI; 2002-144965/19.
DR P-PSDB; ABB08642.
XX
PT New scavenger receptor-like protein for diagnosis, prevention and
PT treatment of autoimmune disease, such as rheumatoid arthritis -
XX
PS Claim 3; Fig 1; 38pp; Japanese.
XX
CC The invention relates to a human scavenger receptor-like protein. The
CC protein is useful as a target molecule for diagnosis, prevention and
CC treatment of autoimmune diseases such as rheumatoid arthritis.
XX
SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;

Alignment Scores:
Pred. No.: 1,74e-58 Length: 2262
Score: 1940.00 Matches: 342
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x ABA97932 (1-2262)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1234 ATGCAACAAGATTGTAGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1293
QY 21 MetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhethrIle 40
Db 1294 ATGGAGAATAAGCTAGTACTCAAGCATGGTGGTCAATCAAGAAATTTACAATA 1353
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlySerGlnGlyPro 60
Db 1354 CTACAGGTGCCACGGGCCCCAGGGTCCAGAGGTGACAGAGGTCCAGGAGCCCCCT 1413
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro 80
Db 1414 GGCCCACTGGCAACAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCCTGGCCCT 1473
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
Db 1474 CGGGGTGACAGAGGCCCAATTGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGA 1533
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
Db 1534 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAG 1593

QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyLeuProGlyPro 140
Db 1594 GCGCCAGTGGGAGCCAGGCCCGCGGCCACACAGGCAAGAGAGGACTCCCGGCCCT 1653
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1654 CAGGGCCCTCCCTGGCTTCCAGGACTTCAGGGCACCCTGGGGAGCTGGGTGGCTGA 1713
QY 161 ProArgGlyLeuProGlyValProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1714 CCTGGGGACTCCAGGCTTGCTGGGTACCAGGATCCAGGCCCAAGGCCCGCCCC 1773
QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1774 GCGCTCTCTGGCCCATCAGAGCGGTGGTGGCTGGCCCTGCAGAAATGAGCAACCCG 1833
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPhethrAspLysCysTrpTyr 220
Db 1834 GCACCGGAGGACAATGGCTGCCCGCTCCTGGAAGAACTTCACAGACAAATGCTACTAT 1893
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1894 TTTTCAGTTGACAAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAACACAGCTCTCA 1953
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1954 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATGATAAAAAACAGATGGTAGG 2013
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2014 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTG 2073
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2074 GATGGACATCTCCAGACTACAAAATTCGAAGCTGGACCGCGGATTAACCTGGGTGAT 2133
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2134 GGCCATGGGCCAGGAGAGACTGTGCTGGTTGATTATGCTGGGCAGTGAACGATTTC 2193
QY 321 GlnCysGluAspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2194 CAATGTGAAGAGCTCAATAACTTCAATTTCGAAAAAGACAGGAGACAGTACTGTCTCT 2253
QY 341 AlaLeu 342
Db 2254 GCATTA 2259

RESULT 3
AA158842
ID AA158842 standard; cDNA; 2318 BP.
XX
AC AA158842;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1045.
XX
KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;
XX
DR WPI: 2001-442253/47.
XX P-PSDB; AAM39686.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1045; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Alignment Scores:
Pred. No.: 1.77e-58 Length: 2318
Score: 1940.00 Matches: 342
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x AAI58842 (1-2318)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValile 20
DB 663 ATGCACACAGATTGATGAGGTGCGAGGTAGACACTGGAAGTACCAACTATCAGTGATT 722

QY 21 MetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrile 40
DB 723 ATGGAAGAAATGAAGCTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTACAATA 782

QY 41 LeuGlnGlyProGlyProGlyProArgGlyAspArgGlySerGlnGlyProPro 60
DB 783 CTACAAGGTCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCGAGGACCCCT 842

QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
DB 843 GCGCCCACTGGCAACAGGAGCAGAAAGGAGAGGGGGAGCCCTGGACCATCTGCCCCCT 902

QY 81 AlaGlyGluArgGlyProTleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
DB 903 GCGGGTGAGAGAGGGCCCAATTTGACACAGCTGTGTCCTCCCGGAGAGCGGTGGCGCAAGGA 962

QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
DB 963 TCTAAAGGCTCCAGAGGCCCAAGAGCTCCGCTGGTTCCTCCCTGGGAAGCCGCGCCCTCAG 1022

QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
DB 1023 GGCCCCAGTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGAGACTTCCCGGCGCT 1082

QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
DB 1083 CAGGGCCCTCCTGGCTTCAGGAGACTTCAGGGCACCCTGGGGAGCGCTGGGTGCGCTGA 1142

QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
DB 1143 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAGGATGCCAGGCCCAAGGCCCGCCC 1202

QY 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
DB 1203 GGCCCTCCTGGCCCATCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCCCAACCCCG 1262

QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
DB 1263 GCACCGGAGGACAATGGCTGCCCGCTCACTGGAAAGAACTTCACACACAATGCTACTAT 1322

QY 221 PheSerValGlyLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
DB 1323 TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGCAAGACAAGTCTTCA 1382

QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
DB 1383 CATCTGTGTTTTCATAAACAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGGTAGGG 1442

QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
DB 1443 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGACCGTGAACCAAGTGAAGTGGCTG 1502

QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
DB 1503 GATGGGACACTCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTCTAT 1562

QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleLysAlaGlyGlnTrpAsnAspPhe 320
DB 1563 GGCCATGGCCAGGAGAAAGCTGTGCTGGGTGATTATGCTGGGCGAGTGGAAACGATTTC 1622

QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
DB 1623 CAATGTGAAGACGTCAATAACTTCAATTCGAAAAAAGACAGGGAGACAGTACTGTCTATCT 1682

QY 341 AlaLeu 342
DB 1683 GCATTA 1688

RESULT 4
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX
AC AAH43036;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRCL-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 74..2302
FT /tag= a
FT /product= "scavenger receptor"
XX

PN WO200159107-A1.
 XX 16-AUG-2001.
 XX 08-FEB-2001; 2001WO-JP00874.
 XX 14-FEB-2000; 2000JP-0035155.
 PR 10-OCT-2000; 2000JP-0309069.
 PA (FUSO) FUSO PHARM IND LTD.
 XX Wakamiya N;
 PI WPI; 2001-497076/54.
 XX P-PSDB; AAG63346.
 DR New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 PT useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation
 XX Claim 2; Page 79-84; 118pp; Japanese.
 XX The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.
 XX
 SQ Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,93e-58 Length: 2628
 Score: 1940.00 Matches: 342
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-09-763-712A-2_COPY_206_547 (1-342) x AAH43036 (1-2628)
 QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
 DB 1274 ATGCAACAAGATTTCATGAGCTCGAGGTTAGACACTGAAGTAGCCAACTTATCATGATT 1333
 QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
 DB 1334 ATGGAAGAAATGAAGCTAGTACTCCCAAGCATGCTCAGCTCATCAAGAAATTTTACAATA 1393
 QY 41 LeuGlnGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 60
 DB 1394 CTACAGGTGTCACCGGGGCCCAAGGGTCCAGAGGTGACAGAGGATCCAGGAGCCCT 1453
 QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGlyLysGlyGluProGlyProGlyPro 80
 DB 1454 GCGCCCACTGGCAACAGGGACAGAAAGAGAGAGGGGGAGCCTGGACCCTGGCCCT 1513
 QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyLysGly 100
 DB 1514 GCGGTGAGAGAGGCCCAATTTGACACGCTGGTCCCGGGAGAGCGTGGCGGCAAGGA 1573
 QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
 DB 1574 TCTAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTGGGAAGCCCGGCCCTCAG 1633
 QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
 DB 1634 GCGCCCACTGGGAGCCCGGGGCCCAAGGCTCCCGTGGTTCCTGGGAAGCCCGGCCCTCAG 1693
 QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160

Db 1694 CAGGGCCCTCCTGGGCTTCAGGGACTTCAGGGACCGTTGGGAGCCTGGGGTCTGGA 1753
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 Db 1754 CCTCGGGGACTGCCAGGCTTGGCTGGGTACCCAGGCATGCCAGGCCCGCCCGCC 1813
 QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
 Db 1814 GGCCCTCCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAATCAGCAACCCG 1873
 QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
 Db 1874 GCACCGGAGGACAATGGCTGCCCGCTCCTCAGGAGAACTTCACAGACAAATGCTACTAT 1933
 QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
 Db 1934 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGACTTTCTGTGACAGACAGCTTCA 1993
 QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
 Db 1994 CATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 2053
 QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
 Db 2054 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCTGAGAAATCAATGGAAGTGGCTG 2113
 QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
 Db 2114 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACCGCGATTAACCTGGGTGAT 2173
 QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
 Db 2174 GGCCATGGGCCAGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGAAACGATTTC 2233
 QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
 Db 2234 CAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAAGACAGGAGAGACTGCTCATCT 2293
 QY 341 AlaLeu 342
 Db 2294 GCATTA 2299
 RESULT 5
 ABL95574
 ID ABL95574 standard; cDNA; 2005 BP.
 XX ABL95574;
 AC ABL95574;
 DT 19-JUL-2002 (first entry)
 XX
 DE Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.
 XX Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
 KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
 KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary;
 KW antiarteriosclerotic; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200208284-A2.
 PD 31-JAN-2002.
 XX
 PF 09-JUL-2001; 2001WO-US21735.
 XX
 PR 20-JUL-2000; 2000US-219556P.
 PR 25-JUL-2000; 2000US-220624P.
 PR 25-JUL-2000; 2000US-220664P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 02-AUG-2000; 2000US-222695P.
 PR 17-AUG-2000; 2000US-0643657.
 PR 23-AUG-2000; 2000WO-US23522.

Qy	321	GlcCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer	340
Db	1438	CAATGTGAAGACGTCATAACTTTCATTGGCAAAAGACAGGAGACAGTACTGTCATCT	1497
Qy	341	AlaLeu 342	
Db	1498	GCATTA 1503	
RESULT 6			
ABL88085			
ID	ABL88085	standard; cDNA; 2005 BP.	
XX	AC	ABL88085;	
XX	16-MAY-2002	(first entry)	
XX	Human	PRO7223 cDNA sequence SEQ ID NO:27.	
XX	Human;	angio genesis; cardiant; cytostatic; antiangiogenic; hypotensive;	
KW	KW	vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;	
KW	KW	gene therapy; cardiovascular disorder; endothelial disorder; cancer;	
KW	KW	angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;	
KW	KW	age-related macular degeneration; arterial restenosis; angina;	
KW	KW	rheumatoid arthritis; myocardial infarction; thrombophlebitis;	
KW	KW	lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;	
KW	KW	wound healing; chromosome mapping; gene mapping; gene; ss.	
OS	Homo sapiens.		
XX	XX		
PN	WO200200690-A2.		
XX	03-JAN-2002.		
PD	20-JUN-2001; 2001WO-US19692.		
XX	23-JUN-2000; 2000US-213637P.		
PR	20-JUL-2000; 2000US-219556P.		
PR	25-JUL-2000; 2000US-220624P.		
PR	25-JUL-2000; 2000US-220664P.		
PR	28-JUL-2000; 2000WO-US20710.		
PR	02-AUG-2000; 2000US-222695P.		
PR	17-AUG-2000; 2000US-0643657.		
PR	23-AUG-2000; 2000WO-US23322.		
PR	24-AUG-2000; 2000WO-US23328.		
PR	07-SEP-2000; 2000US-230978P.		
PR	18-SEP-2000; 2000US-0664610.		
PR	18-SEP-2000; 2000US-0665350.		
PR	24-OCT-2000; 2000US-242922P.		
PR	08-NOV-2000; 2000US-070923P.		
PR	08-NOV-2000; 2000WO-US30952.		
PR	10-NOV-2000; 2000WO-US30873.		
PR	01-DEC-2000; 2000WO-US32678.		
PR	20-DEC-2000; 2000US-0747259.		
PR	20-DEC-2000; 2000WO-US34956.		
PR	22-JAN-2001; 2001US-07767609.		
PR	28-FEB-2001; 2001US-0796498.		
PR	28-FEB-2001; 2001WO-US06520.		
PR	01-MAR-2001; 2001WO-US06666.		
PR	09-MAR-2001; 2001US-0802706.		
PR	14-MAR-2001; 2001US-0808689.		
PR	22-MAR-2001; 2001US-0816744.		
PR	05-APR-2001; 2001US-0828366.		
PR	10-MAY-2001; 2001US-0854208.		
PR	10-MAY-2001; 2001US-0854280.		
PR	23-MAY-2001; 2001US-0866028.		
PR	25-MAY-2001; 2001US-0866034.		
PR	25-MAY-2001; 2001WO-US17092.		
PR	30-MAY-2001; 2001US-0870574.		
PR	30-MAY-2001; 2001WO-US17443.		
PR	01-JUN-2001; 2001WO-US17800.		
XX	(GETH) GENENTECH INC.		
XX			
XX			

Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A; Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W; WPI: 2002-090516/12.
P-PSDB; ABB84830.

One hundred and eighty seven nucleic acids encoding PRO polypeptides, useful in diagnosis and treatment of cardiovascular (e.g. myocardial infarction), endothelial or angiogenic disorders in a mammal -

Claim 2; Fig 27; 565pp; English.

ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to ABB85003. The PRO proteins and polynucleotides have cardiant, cytostatic, antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic activities, and can be used in gene therapy. The PRO polynucleotides, proteins, agonists and antagonists are useful for treating or diagnosing a cardiovascular, endothelial or angiogenic disorder in a mammal, e.g. cardiac hypertrophy, trauma, cancer, age-related macular degeneration, atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour angiogenesis (such as breast carcinoma and liver carcinoma) and wound healing. The PRO polynucleotides have applications in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping. ABL88259 to ABL88267 represent primers and probes used in the exemplification of the present invention.

Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

	Alignment Scores:			
Pred. No.:	2.57e-58	Length:	2005	
Score:	1934.00	Matches:	341	
Percent Similarity:	99.71%	Conservative:	0	
Best Local Similarity:	99.71%	Mismatches:	1	
Query Match:	99.69%	Indels:	0	
DB:	24	Gaps:	0	

US-09-763-712A-2_COPY_206_547 (1-342) x ABL88085 (1-2005)

Qy	1	MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerVallie	20
Db	478	ATGCACCAAGATTGTATGAGGTCGAGGTTAGACACTGAAGTACGCCAACTTAATCAGTGATT	537
Qy	21	MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrile	40
Db	538	ATGGAGAAGATGAAGCTAGTAGACTCCAGCATGGTCAGCTCATCAGAAATTTACAATA	597
Qy	41	LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	60
Db	598	CTACAAGGTCCACGGGCCCCACGGGTCCAAGAGGTGACAGAGGATCCCGAGGACCCT	657
Qy	61	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro	80
Db	658	GGCCCCAACTGGCACAAAGGACAGAAGGAGAGAAGGGGAGCCTGGACACCTGGCCCT	717
Qy	81	AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly	100
Db	718	CGGGGTGAGAGAGCCCAATTGGACACAGCTGTGTCCCCTCCGAGAGCGTGGCGCAAGA	777
Qy	101	SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln	120
Db	778	TCTAAGGCTCCACAGGSCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAG	837
Qy	121	GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro	140
Db	838	GGCCCCAAGTGGGACCCAGGCCCCCGGGGCCACCACAGGCAAGAGGACTCCCGGCCCT	897
Qy	141	GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyCluProGlyValProGly	160
Db	898	CAGGGCCCTCTGCTGTCCAGGACATTCAGGGACCGCTTGGGGAGCCTGGGGTGCTCGA	957
Qy	161	ProArqGlyLeuProGlyLeuProGlyValProGlyYmetProGlyProLysGlyProPro	180

Db	958	CCCCGGGAGCTGCACGGCTTGGCTGGGTGGTACAGGCATGCCAGGCCCCCAAGGGCCCCCCC	1017
Qy	181	GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro	200
Db	1018	GGCCCTCTGGCCCATCAGAGGGGTGGTGGCCCTGGCCCTGCAGATGAGCAACCCCG	1077
Qy	201	AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTyrTyr	220
Db	1078	GCACCGGAGGACAAATAGCTGCCGCTCCTACTGGAAAGAACTTCACAGACAAATGCTACTAT	1137
Qy	221	PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer	240
Db	1138	TTTTCAGTTTCAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAACTCTCA	1197
Qy	241	HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLeLysLysGlnMetValGly	260
Db	1198	CATCTGTGTTTTCATAAACACTAGAGAGNACAGCAATGGATAAAAACAGATGGTAGGG	1257
Qy	261	ArgLysHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu	280
Db	1258	AGAGAGAGCCACTGGATCGCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGCCTG	1317
Qy	281	AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis	300
Db	1318	GATGGGACATCTCCAGACTACAAAAATTGGAAAGCTGGACAGCCGGATTAACATGGGGTCAT	1377
Qy	301	GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe	320
Db	1378	GGCCATGGCCAGGAGAACACATGTCTGGTTGATTTATGCTGGCAGTGGACGATTC	1437
Qy	321	GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer	340
Db	1438	CAATGTGAAGACGTCAATAACTTTCATTTTCGAAAAAGACAGGAGACAGTACTGTCATCT	1497
Qy	341	AlaLeu 342	
Db	1498	GCATTA 1503	
RESULT	7		
ID	AA160628	standard; cDNA; 2181 BP.	
XX	AC	AA160628;	
XX	XX		
DT	XX	22-OCT-2001 (first entry)	
XX	XX		
DE	XX	Human polynucleotide SEQ ID NO 4617.	
XX	XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KW	KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	KW	amyotrophic lateral sclerosis; Sly-drager Syndrome; chemotactic;	
KW	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	KW	leukaemia; ss.	
OS	OS	Homo sapiens.	
XX	XX		
PN	XX	WO200153312-A1.	
XX	XX		
PD	XX	26-JUL-2001.	
XX	XX		
PF	XX	26-DEC-2000; 2000WO-US34263.	
XX	XX		
XX	XX	21-JAN-2000; 2000US-0488725.	
PR	XX	25-APR-2000; 2000US-0552317.	
PR	XX	09-JUL-2000; 2000US-0598042.	
PR	XX	19-JUL-2000; 2000US-0620312.	
PR	XX	03-AUG-2000; 2000US-0653450.	
PR	XX	14-SEP-2000; 2000US-0662191.	
PR	XX	19-OCT-2000; 2000US-0693036.	
XX	XX	29-NOV-2000; 2000US-0727344.	
XX	XX		

Db 1082 CAGGGCCCTCCTGGCTTCAGGAGACTTCAGGCACACCTGTGGGAGCCCTGGGTGCTCGA 1141
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 Db 1142 CTCGGGGACTCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 1201
 QY 181 GlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
 Db 1202 GGCCCTCCTGGCCATCAGGAGCGTGTGGCTGGCCCTGAGCAATGAGCAACCCCG 1261
 QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
 Db 1262 GCACCGAGGACAATAGCTGCCCGCTCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 1321
 QY 221 PheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
 Db 1322 TTTTCAGTTGAGAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAAGTCTTCA 1381
 QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
 Db 1382 CATCTGTGTTTATAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGCTAGGG 1441
 QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
 Db 1442 AGAGAGCCACTGGATCGGCTCAGACTCAGAGCGGTGAATGGAAGTGGCTG 1501
 QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
 Db 1502 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCA 1561
 QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
 Db 1562 GGCCATGGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGAAACGATTTC 1621
 QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
 Db 1622 CAATGTGAAGAGCTCAATAACTTCAATTTCCGAAAAGACAGGGGAGACAGTACTGTCT 1681
 QY 341 AlaLeu 342
 Db 1682 GCATTA 1687
 RESULT 8
 AAC66903
 ID AAC66903 standard; cDNA; 2641 BP.
 AC
 AC AAC66903;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
 XX
 KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 KW inflammation; reproductive disorder; cardiovascular disorder;
 KW immune disorder; musculoskeletal disorder; developmental disorder;
 KW gastrointestinal disorder; cell proliferation disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200068380-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12811.
 XX
 PR 11-MAY-1999; 99US-0133643.
 XX
 PR 23-AUG-1999; 99US-0150409.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PA Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 PI Azimzal Y;
 XX

DR WPI; 2001-007395/01.
 DR P-PSDB; AAB27236.

XX Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX
 PS Claim 4; Page 121-122; 129pp; English.

XX The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADS).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.

XX Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;

Alignment Scores:
 Pred. No.: 3,11e-58 Length: 2641
 Score: 1934.00 Matches: 341
 Percent Similarity: 99.71% Conservative: 0
 Best Local Similarity: 99.71% Mismatches: 1
 Query Match: 99.69% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AAC66903 (1-2641)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
 Db 912 ATGCACACAGATTTGATGAGTTCGAGGTAGACACTGAAGTACCACTTATCATGAT 971
 QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLysAsnPheThrIle 40
 Db 972 ATGGAAGAAATGAAGCTAGTACTCAAGCATGGTGCAGCTCATCAAGAAATTTTACAATA 1031
 QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
 Db 1032 CTACAAAGTTCACCGGGCCCCAGGGGTCCAGAGGTGACAGAGATCCACAGGACCCCT 1091
 QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
 Db 1092 GGGCCCACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 1151
 QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
 Db 1152 GGGGTGTAGAGAGGCCCAATTTGACAGCTGCTCCCCCGGAGAGCGTGGCGCAAGGA 1211
 QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro 120
 Db 1212 TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGTTCCTCGTGGGAAGCCCGGCCCTCAG 1271
 QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGlyGluLeuProGlyPro 140
 Db 1272 GGGCCCACTGGGGACCCAGGCCCCCCCGGCCCCACAGGCAAGAGAGGACTCCCGGCCCT 1331
 QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
 Db 1332 CAGGGCCCTCTCTGGCTTCCAGGACTTCAGGGCACCGTTGGGGAGCGCTGGGTGCTGGA 1391
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 Db 1392 CCTCGGGGACTGCCAGGCTTGCCTGGGTACAGGATGCCAGGATCCAGGCCCCCAAGGCC 1451
 QY 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
 Db 1452 GGCCCTCTCTGGGCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGNATGAGCAACCCGG 1511
 QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyr 220

Db 1512 GCACCGGAGCAATAAGCTGCCGCTCCTGGAAGAACTTCACAGACAAATGCTACTAT 1571
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1572 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTGTGAACAGAGCTTCA 1631
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1632 CATCTGTTTTCATAAACACTAGAGAGGAACGAATGATGATAAAAAACAGATGTTAGG 1691
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1692 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGGAAGTGG 1751
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1752 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCCGATAACTGGGGTCA 1811
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1812 GCCCATGGCCAGGAGAACTGCTGGTGTGATTTATGCTGGCAGTGGAACTTTC 1871
QY 321 GlnCysGluAspValAsnAsnPhelleCysGluLysAspArgGluThrValLeuSerSer 340
Db 1872 CAATGTGAAGAGCTCAATAACTTCATTTTCGAAAAAGACAGGAGACAGTACTGTCT 1931
QY 341 AlaLeu 342
Db 1932 GCATTA 1937
RESULT 9
ID ABQ92072
XX ABQ92072 standard; cDNA; 2929 BP.
AC ABQ92072;
DT ABQ92072;
DE 04-OCT-2002 (first entry)
DE Human polynucleotide SEQ ID NO 69.
KW Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antitumor; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
OS Homo sapiens.
XX US2002065394-A1.
XX 30-MAY-2002.
XX 22-DEC-2000; 2000US-0745763.
XX 18-MAR-1998; 98US-0040963.
PA (JACO/) JACOBS K.
PA (MCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (SPAU/) SPAULDING V.
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;

XX WPI; 2002-582343/62.
DR P-PSDB; ABP61859.
XX Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT -
XX Claim 250; Page 231-232; 284pp; English.
XX The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumors. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems, (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polynucleotide of the invention.
XX SQ Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;
Alignment Scores:
Pred. No.: 3 33e-58 Length: 2929
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 24 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x ABQ92072 (1-2929)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1266 ATGCAACAAGATTGTGATGAGGTGCGAGTTAGACACTGGAAGTAGCCAACTATCAGTGATT 1325
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlnLeuLysAsnPhenIle 40
Db 1326 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGTGTCACTCATCAAGAAATTTACAATA 1385
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1386 CTACAAGGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCCGGACCCCT 1445
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 1446 GCGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGGCGCTGGACCACTGGCCCT 1505
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlyGlyGlyGly 100
|||||


```

Db 1506 GCGGTGAGAGGCCCCAAATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAAAGGA 1565
Qy 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1566 TCTAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTCCCTGGGAAGCCCGGCGCTCAG 1625
Qy 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGlyLeuProGlyPro 140
Db 1626 GGCCCCAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGTCCCGCGGCCCT 1685
Qy 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyLeuProGlyValProGly 160
Db 1686 CAGGCCCCCTCTGGCTCCAGGGACTTCAGGGACCCGTTGGGGAGCCCTGGGGTGCCTGGA 1745
Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1746 CCTGGGGACTGCCAGGCTTCCCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 1805
Qy 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1806 GGCCCTCTGCGCCATCAGGAGCGGTGTCGCCCTGGCCCTGCAGATGAGCAACCCCG 1865
Qy 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr 220
Db 1866 GCACGGAGGACAATAGCTGCCCGCCCTCACTGGAAGAACTTCACAGACAAAATGCTACTAT 1925
Qy 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1926 TTTTCAGTTGAGAAGAATTTTGGAGATGCAAAAGCTTTCTGTGAAGACAAAGTCTCA 1985
Qy 241 HisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1986 CATCTTGTTTTCATAAACACATAGAGAGGACAGCAATGGATAAAAAACAGATGGTAGGG 2045
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2046 AGAGAGACCACTGATCGGCTCACAGACTCAGAGCGTGAATGAAATGGAAGTGGGTG 2105
Qy 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2106 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGCGATACCTGGGTCAT 2165
Qy 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2166 GGCCATGGCCAGGAGAGACTGTCTGGTGTGATTATGCTGGCAGTGGAAACGATTTC 2225
Qy 321 GlnCysGluAspValAsnAspPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2226 CAATGTGAAGACGTCATAACTTCATTTTCGAAAAAGACAGGGAGACAGTACTGTCTCT 2285
Qy 341 AlaLeu 342
Db 2286 GCATTA 2291

RESULT 10
AAV55746
ID AAV55746 standard; cDNA; 2930 BP.
XX
AC AAV55746;
XX
DT 23-MAR-1999 (first entry)
XX
DE Human secreted protein clone bv227_1 coding sequence.
XX
KW Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor; ds.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers

```

```

FT CDS 67..693
FT /*tag= a
XX
XX WO9855614-A2.
XX
XX 10-DEC-1998.
XX
XX 01-JUN-1998; 98WO-US11210.
XX
XX 29-MAY-1998; 98US-0087255.
XX 04-JUN-1997; 97US-0886896.
XX 04-JUN-1997; 97US-0886897.
XX 04-JUN-1997; 97US-0886898.
XX 04-JUN-1997; 97US-0886899.
XX 04-JUN-1997; 97US-0886900.
XX 04-JUN-1997; 97US-0886911.
XX 04-JUN-1997; 97US-0886912.
XX 04-JUN-1997; 97US-0886913.
XX 04-JUN-1997; 97US-0886914.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
XX McCoy JM, Racie LA, Spaulding V, Treacy M;
XX
XX WPI: 1999-059912/05.
XX P-PSDB; AAW73628.
XX
XX New polynucleotides encoding secreted human proteins - derived from
XX human foetal brain, adult brain, foetal kidney, adult ovary, adult
XX retina, adult placenta or adult uterus cDNA libraries
XX
XX Claim 26; Page 87-88; 127pp; English.
XX
XX This sequence encodes a human secreted protein of the invention.
XX This DNA sequence was isolated from a human adult brain cDNA
XX library, and was designated clone bv227_1. The DNAs and proteins
XX are predicted to have biological activities which would make them
XX suitable for treating, preventing or ameliorating medical conditions in
XX humans and animals, although no supporting data is given. Suggested
XX activities include nutritional sources or supplements, immune
XX stimulating or suppressing activity, haematopoiesis regulating activity,
XX tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
XX receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
XX invasion suppressor activity, and tumour inhibition activity. The DNAs
XX are also stated to be useful for gene therapy. A host cell transfected
XX with the DNA, or its subfragments and variants is useful for recombinant
XX production of the human secreted protein clones.
XX
XX Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;
XX
Alignment Scores:
Pred. No.: 3.33e-58 Length: 2930
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 20 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AAV55746 (1-2930)
Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1266 ATGCAACAGATTTTCATGAGTTCGAGGTAGACACTGAAGTAGCCCAACTATTCAGTGAT 1325
Qy 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 1326 ATGGAAGAAATGAACCTAGTAGACTCCCAAGCATGGTCAGCTCATCAAGAAATTTTCAATA 1385
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60

```

Db 1386 CTACAGAGTCCACCGGGCCCCAGGGGTCCACAGAGGTGACAGAGGATCCACGAGACCCCT 1445
 QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
 Db 1446 GGGCAACTGGCAACAGGAGCAGAAAGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCT 1505
 QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
 Db 1506 GCGGGTGAGAGAGGCCAATTTGACACAGCTGCTCCCGCCGAGAGAGCGTGGCGCAAGGA 1565
 QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
 Db 1566 TCTAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTTCCTCCGGAAGCCCGCCCTCAG 1625
 QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGly 140
 Db 1626 GGGCCAGTGGGACCCAGGCCCCCGGGCCACCAGGCAAGAGGAGCTCCCGGCCCT 1685
 QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
 Db 1686 CAGGGCCCTCGCTTCCAGGAGCTTCCAGGACCGTTGGGAGCGCTGGGTGCGTGA 1745
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyPro 180
 Db 1746 CCTCGGGGAGCTCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCCC 1805
 QY 181 GlyProGlyProSerGlyAlaValProLeuAlaGlnAsnGluProThrPro 200
 Db 1806 GGGCCCTCGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAACCCG 1865
 QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPhetAspLysCysTyr 220
 Db 1866 GCACCGAGGAGCAATAGCTGCCCGCTCCTACTGGAAGACTTCACAGACAATGCTACTAT 1925
 QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 240
 Db 1926 TTTTCAGTTGAGAAGAAATTTTGGAGATGCAAGCTTTCTGTGAACAAGTCTTCA 1985
 QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
 Db 1986 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGATGATAAAAAACAGATGTTAGG 2045
 QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
 Db 2046 AGAGAGAGCCATGGATCGCCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTG 2105
 QY 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
 Db 2106 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAT 2165
 QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
 Db 2166 GGGCATGGGCCAGGAGAGACTGTGCTGGGTGATTATGCTGGGAGTGAACGATTTC 2225
 QY 321 GlnCysGluAspValAsnAsnPhelIleCysGlyLysAspArgGluThrValLeuSerSer 340
 Db 2226 CAATGTGAAGAGCTCAATAAATTCATTTCCGAAAAAGACAGGGAGAGACGTACTGTCTATCT 2285
 QY 341 AlaLeu 342
 Db 2286 GCATTA 2291
 RESULT 11
 ID AAH43037
 XX AAH43037 standard; cDNA; 2637 BP.
 AC AAH43037;
 XX
 DT 15-OCT-2001 (first entry)
 DE Nucleotide sequence of a human scavenger receptor.
 XX
 KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;

KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 CDS 92..2320
 FT /*tag= a
 ET /product= "scavenger receptor"
 XX
 PN WO200159107-A1.
 XX
 PD 16-AUG-2001.
 XX
 PF 08-FEB-2001; 2001WO-JP00874.
 XX
 PR 14-FEB-2000; 2000JP-0035155.
 PR 10-OCT-2000; 2000JP-0309068.
 XX
 PA (FUSO) FUSO PHARM IND LTD.
 XX
 PI Wakamiya N;
 XX
 DR WPI; 2001-497076/54.
 DR P-PSDB; AAG63347.
 XX
 PS Claim 6; Page 88-93; 118pp; Japanese.
 CC The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.
 XX
 SQ Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;

Alignment Scores:
 Pred. No.: 1,04e-53 Length: 2637
 Score: 1801.00 Matches: 313
 Percent Similarity: 95.91% Conservative: 15
 Best Local Similarity: 91.52% Mismatches: 14
 Query Match: 92.84% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AAH43037 (1-2637)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
 Db 1292 ATGCGACAGACATGATGAGGTCAAGTTAGACTGAAGTGCACCACTTATCAGTGGT 1351
 QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLysAsnPhetThrIle 40
 Db 1352 ATGGAAGAGATGAACCTGGTGTGACTCCAAGCAGGTCACCTATCAAGAACCTTACCATT 1411
 QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
 Db 1412 CTACAAGGTCTCTCTGGCCCCAGAGGTCCAAAAGGTGACAGAGGATCTCAGGACCACT 1471
 QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
 Db 1472 GGTCCAACCTGGCAACAAGGACAGAAAGAGAGAGAGCCCTGGTCCACCTGGCCCT 1531
 QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
 Db 1532 GCGGGTGAGAGGGGCACAATTTGACCAGTCAGTGGCCCTCTCTGGAGAGCGGTGGCAGCAAGGA 1591


```

Db 397 CCGGGCCCTCAGGGCTCCAGTGGGGACCCAGGCCCCCGGCCCCACAGGCAAGAGGGA 456
QY 137 LeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 156
Db 457 CTCGCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCT 516
QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro 176
Db 517 GGGGTGCTGGACCTCGGGGACTGCCAGGCTTGCTGGGTACCCAGGATGCCAGGCCCC 576
QY 177 LysGlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsn 196
Db 577 AAGGGCCCCCGGCCCTCTGGCCCATCAGGAGCGGTGGTGCCTGGCCCTGCGAGAAT 636
QY 197 GluProThrProAlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAsp 216
Db 637 GAGCCAAACCCAGCACCGGAGGACATGGCTGCCGCCCTCAGTGGGAAGAACTTCACAGAC 696
QY 217 LysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGlu 236
Db 697 AAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAA 756
QY 237 AspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLys 256
Db 757 GACAAGCTCTCACATCTTTTTCATAAACCACATAGAGAGGAACAGCAATGGATAAAAAAA 816
QY 257 GlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGlu 276
Db 817 CAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCCAGACTCAGAGCTCAGAGCGTGAATGAA 876
QY 277 TrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAsp 296
Db 877 TGGAACTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGGAT 936
QY 297 AsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTrpAlaGlyGln 316
Db 937 AACTGGGGTCATGGCCATGGGCCAGAGAGAGACTGTGCTGGGTGATTTATGCTGGGCAG 996
QY 317 TrpAsnAspPheGlnCysGluAspValAsnAsnAsnPheIleCysGluLysAspArgGluThr 336
Db 997 TGGACAGATTTCCAAATGTGAAGACGTCATACTTCACTTGGCAAAAAGACAGGGAGACA 1056
QY 337 ValLeuSer 339
Db 1057 GTCTTCACG 1065

RESULT 13
AAH43054
ID AAH43054 standard; DNA; 2256 BP.
XX
AC AAH43054;
XX
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRCL-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT 74..1936
FT /*tag= a
FT /product= "scavenger receptor"
XX
XX WO200159107-A1.
XX
XX
PD 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
XX

```

```

XX
PR 14-FEB-2000; 2000JP-0035155.
PR 10-OCT-2000; 2000JP-0309068.
XX
PA (FUSO ) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
XX WPI; 2001-497076/54.
DR P-PSDB; AAG63350.
XX
XX New Scavenger receptor proteins SRCL-PI with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation
XX
PS Claim 4; Page 105-109; 118pp; Japanese.
XX
CC The present sequence encodes a human scavenger receptor, designated
CC SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Alignment Scores:
Pred. No.: 7.12e-32 Length: 2256
Score: 1158.00 Matches: 218
Percent Similarity: 63.74% Conservative: 0
Best Local Similarity: 63.74% Mismatches: 0
Query Match: 59.69% Indels: 124
DB: 22 Gaps: 1

US-09-763-712a-2_COPY_206_547 (1-342) x AAH43054 (1-2256)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1274 ATGCAACACAGATTGATGAGTGCAGGTAGACACTGAAGTACCACTATCAGTGATT 1333
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 1334 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGTCATCAAGAATTTTACAATA 1393
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlySerGlnGlyProPro 60
Db 1394 CTACAAGGTCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCCAGGACCCCT 1453
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 1454 GGCCCAACTGGCAACAGGGACAGAAAGAGAGAGAGAGAGAGGGGAGCCCTGGACCCCT 1513
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
Db 1514 GCG- 1516
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1516 1516
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
Db 1516 1516
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1516 1516
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180

```

```
Db 1516 ----- 1516
Qy 181 GlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1516 ----- 1516
Qy 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
Db 1517 -----GGCTGCCCGCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1561
Qy 221 PheSerValGluLysGluPheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1562 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTCTGTGAAGACAAAGCTTCA 1621
Qy 241 HisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1622 CATCTGTTTTCATAAACACATAGAGAGGACAGCAATGGATATAAAACAGATGTAGGG 1681
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1682 AGAGAGAGCCACTGGTGGCCTCAGACTCAGAGCTCAGAGCGTGAAATGAATGGAAGTGGCTG 1741
Qy 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1742 GATGGACATCTCCAGACTACAAAATTGGAAGCTGGACAGCCGGATAAATGGGGTTCAT 1801
Qy 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1802 GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGGAACATTTC 1861
Qy 321 GlnCysGluAspValAsnAsnPheIleCysGluLysGluLysAspArgGluThrValLeuSerSer 340
Db 1862 CAATGTGAAGAGCTCAATAACTTATTTCGGAATAAAGACAGGAGACAGTACTGTCTCT 1921
Qy 341 AlaLeu 342
Db 1922 GCATTA 1927

RESULT 14
AAX27858
ID AAX27858 standard; DNA; 3685 BP.
XX
AC AAX27858;
XX
DT 02-JUN-1999 (first entry)
XX
DE Human CSR3 protein coding sequence.
XX
KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX
OS Homo sapiens.
XX
PN W09909159-A1.
XX
PD 25-FEB-1999.
XX
PF 12-AUG-1998; 98WO-JP03602.
XX
PR 30-JUL-1998; 98JP-0230121.
XX
PR 13-AUG-1997; 97JP-0233396.
XX
PA (NIRS ) JAPAN TOBACCO INC.
XX
PI Nakamura Y, Tokino T;
XX
WPI; 1999-181032/15.
DR P-PSDB; AAY00994.
XX
PT Scavenger receptor proteins - for treatment and diagnosis of
XX disorders involving cell stress
```

```
PS Claim 6; Page 142-150; 175pp; Japanese.
XX
CC This sequence encodes the human cellular stress response 3 (CSR3) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
SQ Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other;
```

Alignment Scores:

Pred. No.:	1.07e-10	Length:	3685
Score:	540.00	Matches:	98
Percent Similarity:	66.14%	Conservative:	27
Best Local Similarity:	51.85%	Mismatches:	46
Query Match:	27.84%	Indels:	18
DB:	20	Gaps:	2

US-09-763-712A-2_COPY_206_547 (1-342) x AAX27858 (1-3685)

```
Qy 5 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMet 24
Db 1405 CTGCTCAGTCCCGCGCTGGACCTCAACGCTCCGGAACCTCTCCATGATCGTGAGGAGATG 1464
Qy 25 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro 44
Db 1465 AAGGAGTGGACACACAGCATGGAGAAATCTTCGCAATGTCCACATCTCAGAGGTGCC 1524
Qy 45 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGly 64
Db 1525 CCCGCCCTCCAGGACCAAGAGATTCAAGAGAGATATGGCGCTGAAGGCGCTGTGGC 1584
Qy 65 AsnLysGlyGlnLysGlyGlu-----Lys 72
Db 1585 GGCAGAGGCCGAAAGGAGAGACCCCGGCATCTTTGGGCCCTGGGACCCAGGCTCTCTCAG 1644
Qy 73 GlyGluProGlyProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro 92
Db 1645 GGGCAACCTGGAGAGCGCGGCTGTGGAGAAAGGGGCCCTGTGGCCCTCGAGGGTTC 1704
Qy 93 ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly 112
Db 1705 CCAGGCCTCAAAGGCTCAAAGGCGAGCTTTGGAACCTGGAGGCGCAGAG----- 1752
Qy 113 SerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProPro 132
Db 1753 -----GGACAGCCAGGCCCAAAAGGGGACATAGGGCCCCCAGGGCCAGAGGCCCCCG 1806
Qy 133 GlyLysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThr 152
Db 1807 GGGTCTCCAGGGCCCTCAGGGCTCAGGGAACCGGGAATTCAGGGGAAGACAGGGTCA 1866
Qy 153 ValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 172
Db 1867 CCAGGCCAGCGGGGCCCATGGGCCCTAAGGTGAACACGAGGATCCAGGGTCCCCCTGCT 1926
Qy 173 MetProGlyProLysGlyProProGly 181
Db 1927 CTCCGGGGCCCTCCAGGTCACACAGGA 1953

RESULT 15
AAX27856
ID AAX27856 standard; DNA; 3810 BP.
XX
AC AAX27856;
XX
DT 02-JUN-1999 (first entry)
XX
DE Human CSR1 protein coding sequence.
XX
KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
```

KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX
OS Homo sapiens.
PN WO9909159-A1.
XX
PD 25-FEB-1999.
XX
PF 12-AUG-1998; 98WO-JP03602.
XX
PR 30-JUL-1998; 98JP-0230121.
PR 13-AUG-1997; 97JP-0233396.
XX
PA (NISR) JAPAN TOBACCO INC.
XX
PI Nakamura Y, Tokino T;
XX
DR WPI; 1999-181032/15.
DR P-PSDB; AAY00992.
XX
PT Scavenger receptor proteins - for treatment and diagnosis of
PT disorders involving cell stress
XX
PS Claim 4; Page 119-127; 175pp; Japanese.
XX
CC This sequence encodes the human cellular stress response 1 (CSR1) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
SQ Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;

Alignment Scores:

Pred. No.:	1.1e-10	Length:	3810
Score:	540.00	Matches:	98
Percent Similarity:	66.14%	Conservative:	27
Best Local Similarity:	51.85%	Mismatches:	46
Query Match:	27.84%	Indels:	18
DB:	20	Gaps:	2

US-09-763-712a-2_copy_206_547 (1-342) x AAX27856 (1-3810)

QY	5	LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMet	24
Db	1530	CTGCTCAGTGCCTCCGGCTGGACCTCAACGTCGGGAACCTCTCCATGATCGTGAGGAGATG	1589
QY	25	LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPherThrIleLeuGlnGlyPro	44
Db	1590	AAGGCAGTGGACACACAGCATGGAGAAATCCTTCGCAATGTCACCATCTACGAGGTGCC	1649
QY	45	ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGly	64
Db	1650	CCCGGCCCTCCAGGACCAAGAGGATTCAAAGGAGATATGGCGCTGAAAGGGCTGTGGC	1709
QY	65	AsnLysGlyGlnLysGlyGlu-----Lys	72
Db	1710	GGCAGAGGGCCGAAAGAGGACCCCGCATCTTGGGGCCCTCGGACCCAGGGTCCTCAG	1769
QY	73	GlyGluProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro	92
Db	1770	GGCAACCTGGAGAGCCCGGCTGTGGGAGAAAGGGCCCTGTGGCCCTCGAGGGTTC	1829
QY	93	ProGlyGluArgGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly	112
Db	1830	CCAGGCCCTCAAAGGCTCAAAGGGCAGCTTTGGAAGTGGAGGGCCGAGA-----	1877
QY	113	SerProGlyLysProGlnGlyProSerGlyAspProGlyProProGlyPro	132
Db	1878	-----GGACGCCAGGCCCAAAAGGGGACATAGGCCCCCAGGGCCCAAGGGCCCCCG	1931

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:26:21 ; Search time 36.3343 Seconds
(without alignments)
2886.622 Million cell updates/sec

Title: US-09-763-712A-2_COPY_206_547

Perfect score: 1940

Sequence: 1 MQQDLMSRLDTEVANSVI.....EDVNFICEKDRVTLSSAL 342

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09763712/runat_14032003_100951_18187/app_query.fasta_1.1877
-DB-Issued_Patents_NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=Dits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn_1.1.108 @runat_14032003_100951_18187 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	459.5	23.7	810	1	Sequence 60, Appl
2	452.5	23.3	3181	1	Sequence 1, Appli
3	448.5	23.1	756	1	Sequence 50, Appl
4	436.5	22.5	1608	4	Sequence 19, Appl
5	428	22.1	4031	1	Sequence 1, Appli
6	424.5	21.9	1572	4	Sequence 39, Appl
7	419	21.6	2543	1	Sequence 11, Appl
8	419	21.6	2543	3	Sequence 11, Appl
9	412.5	21.3	432	1	Sequence 48, Appl
10	409	21.1	3394	1	Sequence 4, Appli
11	402.5	20.7	1416	1	Sequence 1, Appli
12	402.5	20.7	1416	2	Sequence 1, Appli

Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 20, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 40, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 264, App
Sequence 62, Appli
Sequence 9, Appli
Sequence 9, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 372, App
Sequence 137, App
Sequence 94, Appli
Sequence 9, Appli
Sequence 66, Appli
Sequence 116, App
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 50, Appli

13 402.5 20.7 1416 3 US-09-167-364-1
14 402.5 20.7 1416 4 US-09-439-897-1
15 400.5 20.6 4359 4 US-09-484-970B-4
16 399 20.6 1881 4 US-09-029-348-20
17 394.5 20.3 1560 2 US-08-794-795-5
18 394.5 20.3 1560 4 US-09-249-200-5
19 394 20.3 5102 1 US-08-494-168-1
20 393.5 20.3 503 4 US-09-297-269-40
21 389.5 20.1 1703 2 US-08-794-795-1
22 389.5 20.1 1703 4 US-09-249-200-1
23 380.5 19.6 1868 1 US-08-392-367B-1
24 380.5 19.6 1868 3 US-08-893-467A-1
25 370 19.1 1560 4 US-09-453-702B-264
26 370 19.1 61663 4 US-09-453-702B-62
27 355.5 18.3 2383 4 US-09-523-487-9
28 355.5 18.3 2409 3 US-09-320-095-9
29 350 18.0 1458 4 US-09-111-470-3
30 344.5 17.8 1839 1 US-08-383-744-1
31 344.5 17.8 1839 2 US-08-999-336-1
32 344.5 17.8 1839 5 PCT-US96-01427-1
33 330.5 17.0 392 4 US-09-404-879A-372
34 325.5 16.8 48908 4 US-09-453-702B-137
35 323.5 16.7 3404 4 US-09-453-702B-94
36 322 16.6 1370 4 US-09-111-470-9
37 321.5 16.6 9827 4 US-09-453-702B-66
38 320.5 16.5 45175 4 US-09-453-702B-116
39 316.5 16.3 2824 2 US-09-010-928B-3
40 315.5 16.3 2851 4 US-09-535-521-1
41 315.5 16.3 2851 4 US-09-535-521-3
42 310.5 16.0 876 4 US-09-535-521-4
43 310.5 16.0 876 4 US-09-535-521-6
44 306 15.8 18609 4 US-08-943-731-1
45 305.5 15.7 756 1 US-08-642-255-50

ALIGNMENTS

RESULT 1
US-08-642-255-60
; Sequence 60, Application US/08642255
; Patent No 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"
US-08-642-255-60

Alignment Scores:
Pred. No.: 1,74e-17 Length: 810
Score: 459.50 Matches: 93
Percent Similarity: 53.16% Conservative: 8
Best Local Similarity: 48.95% Mismatches: 66
Query Match: 23.69% Indels: 23
DB: 1 Gaps: 4

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-642-255-60 (1-810)

QY 42 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 61
Db 52 AAAGGTGCTCCGGGACCTGCAGGCCACCGGTAGCGTGGCGATCCGGGACACCGGT 111
QY 62 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProProGlyProAla 81
Db 112 GCACCTGGCCCGGACCGGTCCGCTGGATCTAGAGGTGACCCAGGACCGCTGCTCCG 171
QY 82 GlyGluArgGlyPro-----LysGlyPro 89
Db 172 GGACCTGCAGGCCACCGGTGATCGGTGCGATCCGGGACCGGTGACCTGGCCCA 231
QY 90 AlaGlyProProGlyGluArgGlyGly-----LysGlySerLysGlySerGlnGly 106
Db 232 CGCGGTCCGCTGGATCTAGAGGTGACCCAGGACCGCTGGTCTCCGGGACCTGCAGGC 291
QY 107 ProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspPro 126
Db 292 CCACCGGTAGCGTGGCGATCCGGGACCGGTGACCTGGCGGACCGGTGCGCCT 351
QY 127 GlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProProGlyPhe 146
Db 352 GGATCTAGAGTGCACCGGACCGGTGCTGCTCCGGGACCTGCAGGCCACCGGTAGC 411
QY 147 GlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGly 166
Db 412 CGTGGCGATCCGGGACCGGTGACCTGGCGGACCGGTGACCTGGCGGACCGGTAGAGT 471
QY 167 LeuProGlyValProGlyMetProGlyProLysGlyPro-----Pro 180
Db 472 GACCGAGGACCGCTGGTCTCCGGGACCTGCAGGCCACCGGTGACCGGTGCGGATCCG 531
QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 532 GGACACCGGTCACCTGGCCCGGCGGTTCGCCCTGGATCTAGAGGTGACCGGACCGG 591
QY 201 -----AlaProGluAsnGlyCysPro 208
Db 592 CTGGTGTCCGGGACCTGCAGGCCACCG 621

RESULT 2

US-08-655-086-1
Sequence 1, Application US/08655086
Patent No. 5821089
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOTT A.
APPLICANT: BUECHTER, DOUGLAS
APPLICANT: ZHANG, GUANGHUI
APPLICANT: CONNOLLY, KEVIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVERTON BOULEVARD
CITY: UNIONDALE
STATE: NY

COUNTRY: US
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,086
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S.
REGISTRATION NUMBER: 32,063
REFERENCE/DOCKET NUMBER: 203-1632
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-228-8484
TELEFAX: 516-228-8516

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cdna
US-08-655-086-1

Alignment Scores:
Pred. No.: 1,57e-16 Length: 3181
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 23.32% Indels: 49
DB: 1 Gaps: 8

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-655-086-1 (1-3181)

QY 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 893 GTTCAAGGACCCCTGGCCCTGCTGGAGAGGAAGAAAGGAGGAGCTCGAGGTGAACCC 952
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro----- 75
Db 953 GGACCACTGGCTGCGGACCCCTGGGAGCGGTGGTGGACCTGGTAGCCTGCTTTC 1012
QY 76 -----GlyProGlyProAlaGlyGluArgGlyProLleGly 88
Db 1013 CCTGGCGCAGATGCTGTGCTGCTCCCAAGGTCCTGGTGAAGCTGCTGCTGCTGCTGCT 1072
QY 89 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerGlySerGln 105
Db 1073 CCGCTGGCCCAAGGATCTCTGCTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1132
QY 106 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 125
Db 1133 GTTGCAAGGCTGACTGGAAGCCCTGGCAGCCCTGGTCTGCTGCTGCTGCTGCTGCTGCT 1186
QY 126 ProGlyProProGlyProGlyLysGlyGluLeuProGlyProGlnGlyProPro--- 144
Db 1187 ---GGCCCCCTGGTCCCGCGGTCAAGATGCTGCGCCCGGACCCCGGACCCCTGGT 1243
QY 145 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 156
Db 1244 GCGCGTGGTCAGGCTGGTGTGATGGGATTCCTGACCTAAAGGTGCTGCTGGAGAGCC 1303
QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal----- 170
Db 1304 GCAAGGCTGGAGGAGGAGGTGTCCCGGACCCCTGGGACCTGCTGCTGCTGCTGCTGCT 1363
QY 171 ProGlyMetProGlyProLysGlyProGlyProGlyProGlyProGlyProGlyPro 187
Db 1364 GATGGAGGCTGGAGCTCAGGAGCCCTGGGACCCCTGGGACCTGCTGCTGCTGCTGCTGCT 1423


```

QY 188 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 201
Db 1424 GAACAGGCGCTGCTGGCTCCCGCC---GGATTCCAGGGTCTCGTGGTCTCT 1480
QY 202 -----ProGluAspAsnGlyCysPro 208
Db 1481 CCAGGTGAAGCAGGCAACCTGGTGAACAGGGTGTCTCT 1519

RESULT 3
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A5556-3/BI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-50

Alignment Scores:
Pred. No.: 6,41e-17 Length: 756
Score: 448.50 Matches: 87
Percent Similarity: 55.03% Conservative: 6
Best Local Similarity: 51.48% Mismatches: 63
Query Match: 23.12% Indels: 13
DB: 1 Gaps: 2

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-642-255-50 (1-756)
QY 43 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
Db 10 GGGCCAGCAGGTCCGAGGCGCGGATGCCCCAGAGGCCGGAAGGTGGCTGGACCG 69
QY 63 ThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGly 82
Db 70 GCTGGTCCACCGGGTCTCCGGGACCTGCAGCCCGCCAGGTGGCGCTGGACCGGTGT 129
QY 83 GluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlySerLys 102
Db 130 CCACCGGGTGTCTCCGGGACCTGCAGGCGCGCCAGGTGGCGCTGGACCGGTGTCT 189

```

```

QY 103 GlySerGlnGlyProLysGlySerArgGlySerPro-----GlyLysProGlyPro 119
Db 190 GGTGCTCCGGAGCCTGCAGCGCCGCGAGGTGGCGCTGGACCGGCTGTCCACCGGTGCT 249
QY 120 GlnGlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGluGlyLeuProGly 139
Db 250 CCGGGACCTGCAGCGCCGCGAGGTGGCGCTGGACCGGCTGTCCACCGGTGTCTCCGGA 309
QY 140 ProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 159
Db 310 CTTGAGGCGCCGCGAGGTGGCGCTGGACCGGCTGTCCACCGGCTGTCTCCGGAGCTGCA 369
QY 160 GlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyPro 179
Db 370 GGGCCGCGAGGTGGCGCTGGACCGGCTGTCCACCGGCTGTCCCGGACCTGCAGGCGCG 429
QY 180 ProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThr 199
Db 430 CCAGGTGGCGCTGGACCGGTGT-----CCACCG 459
QY 200 ProAlaProGluAspAsnGlyCysPro 208
Db 460 GGTGCTCCGGAGCCTGCAGGCGCGCA 486

RESULT 4
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-19

Alignment Scores:
Pred. No.: 5,93e-16 Length: 1608
Score: 436.50 Matches: 102
Percent Similarity: 47.06% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 67
Query Match: 22.50% Indels: 70
DB: 4 Gaps: 9

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-029-348-19 (1-1608)
QY 30 LysHisGlyGlnLeuLysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGly 49
Db 59 CAACATGCCAATCTTTACAGAGGAA-CTGTGAAGAGGCGCCCGGAGATAGAGGA 117
QY 50 ProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlnLys 69
Db 118 CCACGTGGAGAAAGGGTCCACCGAGCCCGCCAGGAGAGATGTTGAAGATGTCCACA 177
QY 70 GlyGluLysGlyGluProGlyProGlyPro----- 80
Db 178 GGGCTCTCTGGTCCACCTGTCTCTGGCGCCCTGTCTCGGTGGGAACATTTGTGTCT 237
QY 81 -----AlaGlyGluArgGlyProIleGlyProAla 90
Db 238 CAGTATATGGAAGAGGTGGACTGGCCCTGGACCAATGGGCTTAATGGGACCTAGA 297
QY 91 GlyProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySer 110

```


.....

Best Local Similarity: 29.30% Mismatches: 91
 Query Match: 21.60% Indels: 176
 DB: 1 Gaps: 9

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-555-669-11 (1-2543)

QY 27 ValAspSerLysHisGlyGlnLeuThrLeuLeuLysAsnPheThrIleLeuGlnGlyProGly 46
 DB 986 CTGATGGCCAGAGGAGAGGCTGGTCGCAAC-----GGTCTCCGGGA 1030

QY 47 ProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLys 66
 DB 1031 GAGAGGGCCCCAACGGGCTCCGGGCTCCCTGGACGAGCGGGCTCCAAAGGCGAGAG 1090

QY 67 GlyGlnLysGlyGlnLysGlyGlnProGlyProGlyProGlyProAlaGlyGluArgGlyPro 86
 DB 1091 GAGAACGGGGCAGAGCTGGGAGCTGGTGAGCGCGGCCCTCTGGAGAGCCAGCGCTC 1150

QY 87 IleGlyProAlaGlyProGlyGlnArgGlyGlyGlySerLysGlySer----- 104
 DB 1151 CTGGAGATGCTGGCATGCTGGGAGCGCGGTGAGGCTGGCCACCGGGCTCAGCGGG 1210

QY 105 -----cInGlyProLys 108
 DB 1211 GCCTCGGGCCACAGGCCCTCCCGAGCCCTGGTGTCGAGGCTTCAGGGCCAGAG 1270

QY 109 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 128
 DB 1271 GGCAGATGGAGACCCCGCCCTCCAGGCCCCCGAGGCTCCGAGGTGACGTGGCGCAG 1330

QY 129 ProGlyProGlyLysGlyGlnLeuProGlyProGlnGlyProGlyPheGlnGly 148
 DB 1331 CGGGTCCGGAGGTGCCAAGGCCCTTAAGGGAGACCAGGGTATTGCAGGTTCCGACG 1390

QY 149 LeuGlnGlyThrValGlyGlu----- 155
 DB 1391 CTTCCTGGGGATAAAGGAGAACTGGTGCCACGGCGCTGGTCCGACCCCAAGAGAGT 1450

QY 155 ----- 155
 DB 1451 GGCAGTCGAGGGAGGTGGGGCCCCAAAGGCCACCCAGGGTCCCAAGGCCACGCGGT 1510

QY 156 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 172
 DB 1511 CAGGTTGTCGGGGCCCCCGGCTCTCTGGGCTTCGAGGGGCTCCCGGGTTCCTGTC 1570

QY 173 Met----- 173
 DB 1571 ATCAGGGGAAGCCGGGAGTTCGGGGGAAGAGAGGCCAGCGGCATCAGGGAGCTG 1630

QY 173 ----- 173
 DB 1631 TGTGGGGGATGATCAGCGAACAAATTCACAGTTAGCCGCCACCTAAGGAAGCCTTTG 1690

QY 174 -----ProGlyProLysGlyProGlyProGlyProGlyProGlyProSer 186
 DB 1691 GCACCGGGTCCATGTTGGTGGCGCCGCTCCAGCTGGCCCGCTGGCGCCCGCCAGCCCA 1750

QY 187 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 198
 DB 1751 GGCTCCATGGTTCACCTGGCGCTCGAGGACCCCGGATACCGGGGTCCTCCAGGGAG 1810

QY 199 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs 216
 DB 1811 CTGGGAGACCCCGGCCAGAG----- 1832

QY 216 pLysCysTyTrpPheSerValGluGlnLeuPheGluAspAlaLysLeuPheCys 236
 DB 1833 -----GAAACACGGGTGACAGAGGAGACAAAG----- 1859

QY 236 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 256
 DB 1860 -----CGCGGGCAGGAGCAG----- 1874

QY 256 sglnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG 276
 DB 1875 -----GGCTGGACGGCCTGAAGGAGACCGAGGGCCCCAAGGA 1912

QY 276 uTrpLysTrpLeuAspGlyThrSerProAspTyTrpLysAsnTrpLysAlaGlyGlnProAs 296
 DB 1913 CCCCAGGG-GTGCCCGGCACCAAGCAGCGCAGGACGCGTGTCTCCCGCGAGCCTGG 1971

QY 296 pAsnTrpGlyHisGlyHisGlyProGly----- 305
 DB 1972 GCCTCCCGAGATCTCGGCTTCAGGTCCTATGGGGCCCAAGGACACCGGGGATCTG 2031

QY 306 -----GluAspCysAlaGlyLeuIleTyTrpAlaGly 315
 DB 2032 CGACACCTCAGCCTGCCAAGGAGCGGTGTAGGAGGG 2068

RESULT 8
 US-09-073-663-11
 ; Sequence 11, Application US/09073663
 ; Patent No. 6127523
 ; GENERAL INFORMATION:
 ; APPLICANT: Brewton, Richard G.
 ; APPLICANT: Mayne, Richard
 ; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McGregor & Adler, LLP
 ; STREET: 8011 Candle Lane
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh OS 8.1
 ; SOFTWARE: Microsoft Word for Macintosh
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/073,663
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Benjamin Aaron Adler, Ph.D., J.D.
 ; REGISTRATION NUMBER: 35,423
 ; REFERENCE/DOCKET NUMBER: D5913D
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (713) 777-2321
 ; TELEFAX: (713) 777-6908
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2543 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 47..2098
 ; US-09-073-663-11

Alignment Scores:
 Pred. No.: 8,17e-15 Length: 2543
 Score: 419.00 Matches: 121
 Percent Similarity: 35.59% Conservative: 26
 Best Local Similarity: 29.30% Mismatches: 91
 Query Match: 21.60% Indels: 176
 DB: 3 Gaps: 9

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-073-663-11 (1-2543)

QY 27 ValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProGly 46
 DB 986 CTGATGGCCAGAGGAGAGGCTGGTCGCAAC-----GGTCTCCGGGA 1030

```

Qy 296 pAsnTrpGlyHisGlyHisGlyProGly----- 305
Db 1972 GCCTCCGGAGATCTCTGGGCTTCCAGGTCATTTGGGGCCACAGGGGACACCGGGGATCTG 2031
Qy 306 -----GluAspCysAlaGlyLeuIleTyrAlaGly 315
Db 2032 CGACACTCAGCCTGCCAAGGAGCCGTGTAGGAGGG 2068

RESULT 9
US-08-642-255-48
: Sequence 48, Application US/08642255
: Patent No. 5773249
: GENERAL INFORMATION:
: APPLICANT: CAPPELLO, Joseph
: APPLICANT: FERRARI, Franco A.
: TITLE OF INVENTION: High Molecular Weight Collagen-Like
: TITLE OF INVENTION: Protein Polymers
: NUMBER OF SEQUENCES: 135
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/642,255
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: ROWLAND, Bertram I.
: REGISTRATION NUMBER: 20,015
: REFERENCE/DOCKET NUMBER: A55556-3/BIR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEX: 910 277299 FHT UR
: INFORMATION FOR SEQ ID NO: 48:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 432 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "synthetic"
US-08-642-255-48

Alignment Scores:
Pred. No.: 3,28e-15 Length: 432
Score: 412.50 Matches: 80
Percent Similarity: 55.70% Conservative: 3
Best Local Similarity: 53.69% Mismatches: 57
Query Match: Indels: 9
DB: Gaps: 3

US-09-763-712A-2_COPY_206_547 (1-342) x US-08-642-255-48 (1-432)

Qy 43 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
Db 10 GGCCCAAGAGTCCGAAGGGCGCATGCCAGCAGCCGAAAGGTGCGCTGGACCG 69
Qy 63 ThrGluAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGly 82
Db 70 GCTGGTCCACGGGTGCTCGGGAGCCTGCAGGCCGCCAGGTGCGCTGGACCGCTGGT 129
Qy 83 GluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLys 102

```

Db 130 CCACGGGTCTCCGGGACCTCCAGCCCGCCAGGTGCGCTGGACCGGCTGGTCCACCG 189
Qy 103 GlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 122
Db 190 GGTGCTCCGGGACCT- - - - -GCAGGCCCGCCAGGTGCGCTGGACCGGCTGGTCCA 240
Qy 123 SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGly 142
Db 241 CCGGGTGTCTCCGGGACCTCCAGCCCGCCA- - - - -GGTGGCTGGACCGGTGTGT 291
Qy 143 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 162
Db 292 CCACCG- - - - -GGTCTCCGGACCTGCAGCCCGCCAGGTGCGCTGGACCGGT 342
Qy 163 GlyLeuProGlyLeuProGlyValProGlyMetProGlyLysGlyProGlyProGlyPro 182
Db 343 GTCCACCGGGTGTCTCCGGACCTGCAGCCCGCCAGGTGCGCTGGACCGGTCCAGCCG 402
Qy 183 ProGlyProSerGlyAlaValValPro 191
Db 403 AAGGAGCTCAGGTCCTCCGCGAGGTCCG 429

RESULT 10

US-08-159-784-4
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 530
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4
Alignment Scores:
Pred. No.: 3.75e-14
Score: 409.00
Matches: 86
Percent Similarity: 48.98%
Best Local Similarity: 43.88%
Query Match: 21.08%
Indels: 42
Gaps: 1
Length: 3394
Matches: 86
Conservative: 10
Mismatch: 42
Indels: 58
Gaps: 7

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-159-784-4 (1-3394)
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer- - - - - 56
Db 553 ATGCCCGGCCCGCCAGGACCTCCAGGCCCGCCAGGCGCTCCAGGACTCTCTTTACGAC 612
Qy 57 - - - - -GlnGlyProProGlyProThrGlyAsnLysGly 67
Db 613 AGCAATGTGTTTGTCTGAGTCCAGCCCGCCGCGCTCCAGGATTGCCAGGGAATCAG- - - 669
Qy 68 GlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGlyProIle 87
Db 670 - - - - -GGCCCTCCAGGACCCCAAGGGCCCAAGAGAGAGTG 705
Qy 88 GlyProAlaGlyProGlyGlu- - - - - 95
Db 706 GGCCCGCCCGGACCCAGGCGAGTTTCGGTTTGACTTTCTTCAGAGGAGGCTGNAATG 765
Qy 96 ArgGlyGlyLysGlySerLysGlnGlyProLysGlySerArgGlySer- - - - - 113
Db 766 AAGGGGAGAGGAGACCGAGGTGATGCAGGACAGAAAGGCGAAAGGGGGAGGCCGG 825
Qy 114 - - - - -ProGlyLysProGlyProGlnGlyProSerGly 124
Db 826 GCGCGCGGTTTCTCGGCTCCAGCTGCCCGGGGCGCCGCGCCAGGCGCCAGCTGGC 885
Qy 125 AspProGlyProGlyProGlyLysGlu- - - - -GlyLeuProGly 139
Db 886 TACCCTGGATTCCAGTCCCAAGGAGAGAGACATCCGGGGCCAGCCCGCCACCTGGA 945
Qy 140 ProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 159
Db 946 CCTCAGGACCCCGCGCATC- - -GGCTACGAGGGGCGCAGGCGCTCCCGGCCCGCCA 1002
Qy 160 GlyProArgGlyLeuProGlyLeuProGly- - - - -ValProGly 172
Db 1003 GGCCCCCGAGGCGCCCTTCATTTCTTGCCCTCACAGCAGCATCATCAGCTTCGGGG 1062
Qy 173 MetProGlyProLysGlyProGlyProGlyProGlyProSerGlyAla 188
Db 1063 CCTCGGGCGCCCTTGGGCGCCCTTGCGCCCTCGAACCATGGGCGCC 1110
RESULT 11
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621,091G
; FILING DATE: 11/30/90
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA: No. 5424408 applicable
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard S.
; REGISTRATION NUMBER: 28180
; REFERENCE/DOCKET NUMBER: 900983/RB

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 972-1400
 TELEFAX: (212) 370-1622
 TELEX: 236268
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1416 base pairs
 TYPE: Nucleic acid
 STRANDEDNESS: Double
 TOPOLOGY: Linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Calf
 STRAIN: Unknown
 INDIVIDUAL ISOLATE: Unknown
 DEVELOPMENTAL STAGE: Unknown
 CELL TYPE: Whole kidney
 IMMEDIATE SOURCE:
 LIBRARY: Bovine lens cDNA
 CLONE: KMC15
 POSITION IN GENOME: No. 5424408 known
 CHROMOSOME/SEGMENT: No. 5424408 known
 PUBLICATION INFORMATION: No. 5424408e
 US-07-621-091G-1

Alignment Scores:
 Pred. No.: 3.6e-14 Length: 1416
 Score: 402.50 Matches: 103
 Percent Similarity: 46.62% Conservative: 21
 Best Local Similarity: 38.72% Mismatches: 103
 Query Match: 20.75% Indels: 40
 DB: 1 Gaps: 8

US-09-763-712A-2_COPY_206_547 (1-342) x US-07-621-091G-1 (1-1416)

QY 42 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 61
 Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGAAACCCAGGATGCTGCTCTCTGGA 171
 QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
 Db 172 CCTCCAGGAGGCTCTAGAGGSCATAAAAGGAGACAGGGGTGATGATGATTCAGGCCTG 231
 QY 77 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 96
 Db 232 CAAAGAGGTCCACCTGGAGCTATAGGAGACATGGGGTCCACGGTCTATCCGGAGCACCA 291
 QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
 Db 292 GGTGTCCCGGTGAGGAGATCCAGGAGCCCTGTGGGCCAAAAGGTAAACACGGGAGGATG 351
 QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProGlyLysGluGly 136
 Db 352 AAAGGAGAGAGGGTAATTCAGGATTTCCAGGACCACCTGGACCTCCAGGGCAAAAGTGA 411
 QY 137 -----LeuPro 138
 Db 412 CCANAAGGACCACCTGAGTACGTGGAGAGCCTGGCAGACAGTGAAGATCATCTCCCTTCCA 471
 QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
 Db 472 GGAAGCCAGCCACCTGTTGAGTGGAGAACACCGGATGCAAGAGAACCCGGGCC 531
 QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
 Db 532 CCAGGACCACAGGAGATCCAGGACCCCTGTGGGCCAAAAGGTAAACACGGGAGGATG 591
 QY 179 ProProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 198
 Db 592 CCACCGAAGTCCCTGGACCACTGGAGAAAGGCAACAAAGGTTGTAAAGAGAGACAA 651

QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 214
 Db 652 GGACCA---CCTGGATCCGATGSCCTGCCAGGCTTCAAGGGGAACC-TGGAGACACTGG 707
 QY 215 ThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
 Db 708 ACC---ACCTGGAGCAGGAGGCGATGATGAGGGCTTTGTCTTTACCGGCACAG----- 758
 QY 235 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
 Db 759 ---CCAGACCACAGCAATTCCTCTCCAGAAAGGACAGCCCTCTATAGTGGTT 815
 QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 272
 Db 816 TTCTCTCTCTTTGTATACAAAGAAATGAACAAGCCCATGGACAGGA----- 850
 QY 273 ArgGluAsnGluTrpLys 278
 Db 861 CCTGGGACACTTGGCAG 878
 RESULT 12
 US-08-399-889-1
 ; Sequence 1, Application US/08399889B
 ; Patent No. 5973120
 ; GENERAL INFORMATION:
 ; APPLICANT: Reiders, Stephen T
 ; APPLICANT: Morrison, Karen E
 ; APPLICANT: Hudson, Billy G
 ; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
 ; FILE REFERENCE: 951263A
 ; CURRENT APPLICATION NUMBER: US/08/399,889B
 ; CURRENT FILING DATE: 1995-03-07
 ; EARLIER APPLICATION NUMBER: 07/621091
 ; EARLIER FILING DATE: 1990-11-30
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1416
 ; TYPE: DNA
 ; ORGANISM: Calf
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1416)
 US-08-399-889-1
 Alignment Scores:
 Pred. No.: 3.6e-14 Length: 1416
 Score: 402.50 Matches: 103
 Percent Similarity: 46.62% Conservative: 21
 Best Local Similarity: 38.72% Mismatches: 103
 Query Match: 20.75% Indels: 40
 DB: 8 Gaps: 8
 US-09-763-712A-2_COPY_206_547 (1-342) x US-08-399-889-1 (1-1416)
 QY 42 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 61
 Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGAAACCCAGGATGCTGCTCTCTGGA 171
 QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
 Db 172 CCTCCAGGAGGCTCTAGAGGSCATAAAAGGAGACAGGGGTGATGAGGAGAGCCTG 231
 QY 77 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg 96
 Db 232 CAAAGAGGTCCACCTGGAGCTATAGGAGACATGGGGTCCACGGTCTATCCGGAGCACCA 291
 QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
 Db 292 GGTGTCCCGGTGAGGAGATCCAGGAGCCCTGTGGGCCAAAAGGTAAACACGGGAGGATG 351
 QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProGlyLysGluGly 136
 Db 352 AAAGGAGAGAGGGTAATTCAGGATTTCCAGGACCACCTGGACCTCCAGGGCAAAAGTGA 411
 QY 137 -----LeuPro 138
 Db 412 CCANAAGGACCACCTGAGTACGTGGAGAGCCTGGCAGACAGTGAAGATCATCTCCCTTCCA 471
 QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
 Db 472 GGAAGCCAGCCACCTGTTGAGTGGAGAACACCGGATGCAAGAGAACCCGGGCC 531
 QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
 Db 532 CCAGGACCACAGGAGATCCAGGACCCCTGTGGGCCAAAAGGTAAACACGGGAGGATG 591
 QY 179 ProProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 198
 Db 592 CCACCGAAGTCCCTGGACCACTGGAGAAAGGCAACAAAGGTTGTAAAGAGAGACAA 651

```
Db 352 AAAGGAAGAGGGTAATTCAGGATTTCAGGACCACCTGGACCTCCAGGGCAAAAGTGA 411
QY 137 -----LeuPro 138
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
Db 472 GGAAGCCAGGCGCCACCTGCTCAGCTGGAGAACACAGGATGCAAGAGAACCCGGGCC 531
QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGlyGly 178
Db 532 CCAGGACCACAGGAGATCAGGACCCCTGTGGGCCCAAAAGGTAAACAGGGGAGGATGTT 591
QY 179 ProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 198
Db 592 CCACCAAGGACTCTCGACCACTGGAGAAAAGGCAACAAAGTTGTAAAGAGAGCA 651
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpIleGlnTrpIle 214
Db 652 GGACCA---CCTGGATCCGATGCTGCCAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 215 ThrAspLysCysTyrThrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
Db 708 ACC---ACCTGCAGCAGGGCAGTGTATGAGGGCTTTGTCTTTACCCGGCAGACAG----- 758
QY 235 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
Db 759 ---CCAGACCACAGCAATTCCTCTGTCCAGNAGGAGCAGACCCCTCTATAGTGGTT 815
QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 272
Db 816 TTCTCTCTCTTTGTACAAGGAATGAACAAGCCCATGGACAGGA-----860
QY 273 ArgGluAsnGluTrpLys 278
Db 861 CTGGGAACACTTGGCAG 878
```

RESULT 13

```
US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-09-167-364-1
```

Alignment Scores:

Pred. No.:	3 6e-14	Length:	1416
Score:	402.50	Matches:	103
Percent Similarity:	46.62%	Conservative:	21
Best Local Similarity:	38.72%	Mismatches:	103
Query Match:	20.75%	Indels:	40
DB:	3	Gaps:	8

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-167-364-1 (1-1416)

```
QY 42 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 61
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGAAACCCAGGCATGCTGCTCCTCTGGA 171
QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
Db 172 CTCTCCAGGAGTCTCTGTAGAAAGGCATAAAGAGAGCAAGGGTTGATGGAGAGCCTGGC 231
QY 77 ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg 96
Db 232 CAAAGAGTCCACCTGGAGCTATAGGACATGGGTCCACAGTCTATCCGGAGCACCA 291
QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
Db 292 GGTGTCCCGGTCAGCCAGGCGCCAGAGGTGATCTCTGATTCATGGATTCACGAGCATG 351
QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGly 136
Db 352 AAAGGAAGAAAGGTAAATTCAGGATTTCCAGGACCACCTGGACCTCCAGGGCAAGTGA 411
QY 137 -----LeuPro 138
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
Db 472 GGAAGCCAGGCGCCACCTGCTCAGCTGGAGAACACAGGATGCAAGAGAACCCGGGCC 531
QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
Db 532 CCAGGACCACAGGATCCAGGACCTGTGGGCCCAAAAGGTAAACAGGGGAGGATGTT 591
QY 179 ProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 198
Db 592 CCACCAAGGACTCTCGACCACTGGAGAAAAGGCAACAAAGTTGTAAAGAGAGCA 651
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 214
Db 652 GGACCA---CCTGGATCCGATGCTGCCAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 215 ThrAspLysCysTyrThrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
Db 708 ACC---ACCTGCAGCAGGGCAGTGTATGAGGGCTTTGTCTTTACCCGGCAGACAG----- 758
QY 235 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
Db 759 ---CCAGACCACAGCAATTCCTCTGTCCAGAAAGGAGCAGACCCCTCTATAGTGGTT 815
QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 272
Db 816 TTCTCTCTCTTTGTACAAGGAATGAACAAGCCCATGGACAGGA-----860
QY 273 ArgGluAsnGluTrpLys 278
Db 861 CTGGGAACACTTGGCAG 878
```

RESULT 14

```
US-09-439-897-1
; Sequence 1, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
```



```
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-09-439-897-1

Alignment Scores:
Pred. No.: 3,6e-14 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 20.75% Indels: 40
DB: 4 Gaps: 8

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-439-897-1 (1-1416)
QY 42 GlnGlyProProGlyProArgGlyProArgGlyProArgGlySerGlnGlyProProGly 61
DB 112 AAAGGAGATCGAGGTCACCTGCTCCAGAGGAACCCAGGATCGCTGCTCTCTGGA 171
QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
DB 172 CTTCCAGGCGAGTCCTCTAGAGGCATAAAGAGACAAGGGTTGATGGAGAGCTGCG 231
QY 77 ProProGlyProAlaGlyGluArgGlyProLleGlyProAlaGlyProProGlyGluArg 96
DB 232 CAAGAGGTTCACCTGGAGCTATAGAGACATAGGAGTCCACAGGTCTCCGGGACACCA 291
QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
DB 292 GGTGTCCTCCCGTCACCCAGGGCCAGAGGTGATCTGATCTATGGATTCAGGCGATG 351
QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProGlyLysGluGly 136
DB 352 AAAGGGAAGAGGTAATTCAGGATTCAGGACCATCCAGGACCATCGGACCTCCAGGGCAAAAGTGA 411
QY 137 -----LeuPro 138
DB 412 CCAAAAGGACCACTCGGAGTACGTGGAGAGCTGGCAGAGCTGAAGATCATCTCCCTTCCA 471
QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
DB 472 GGAAGCCCGAGCCACCTGGTTCAGCTGGAGAACACAGGAGTCAGAGGAACCCGGGCC 531
QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
DB 532 CCAGGACCAACAGGAGATCCAGGACCTCTGGGCCAAGAAAGTAAACACGAGGAGATGCT 591
QY 179 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 198
DB 592 CCACCGAAGAACCTCTGGACCACTGGAGAAAAGCAACAAAGGTTGTAAAGGAGAGCAA 651
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTriLysAsnPhe 214
DB 652 GGACCA---CCTGGATCCGATGGCTGCGAGCTTGAAGGGGAAACCC-TGGAGACACTGG 707
QY 215 ThrAspLysCysTyrrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
DB 708 ACC---ACCTGCAGCAGGGCGAGTATGAGGGCTTTGTCTTTACCGGSCACAG----- 758
QY 235 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
DB 759 ---CCAGACCAACAGAAATTCCTCTCTGTCAGAGGGGACAGCGCGCTCTATAGTGGGTT 815
QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTriLleGlyLeuThrAspSerGlu 272
DB 816 TTCTCTTCTTTGTACAGGAATTAACACACCCCATGACAGGA-----860
QY 273 ArgGluAsnGluTrpLys 278
DB 861 CTGGGAACACTTGGCAG 878
```

RESULT 15

US-09-484-970B-4

; Sequence 4, Application US/09484970B

```
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 4359
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 331365 (1553795CBL)
US-09-484-970B-4
```

```
Alignment Scores:
Pred. No.: 1,38e-13 Length: 4359
Score: 400.50 Matches: 90
Percent Similarity: 54.89% Conservative: 11
Best Local Similarity: 48.91% Mismatches: 64
Query Match: 20.64% Indels: 19
DB: 4 Gaps: 7
```

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-484-970B-4 (1-4359)

```
QY 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArg-----54
DB 448 CTGCCAGGACCCAGGCCACAGGTGCCAAAGGCTCCGAGGAATCCAGGCTTCGCA 507
QY 55 GlySerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGlu 74
DB 508 GGAGTGTATGGAGGACACAGGGTCCAGGGCTTCGCCAGGACGACGAGTCTGTAAGGTTTC 567
QY 75 ProGlyProPro-----GlyProAlaGlyGluArgGlyProLleGlyProAlaGly 91
DB 568 CCAGGACCCCGAGGTTTCATAGGACCCCGAGGATCCAAAGGTGCATGGGCTCCCTGGC 627
QY 92 ProProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArg 111
DB 628 CCAGATGGATGCCAGGTCCCATCGGCTGCCAGGCGCCAGATGGCGCCCTCGGGAAGG 687
QY 112 GlySerProGlyLysPro---GlyProGln---GlyProSerGlyAspProGlyProPro 129
DB 688 GGCCTCCCTGGAGAGTCTCTGGAGCTCAGCCCGGCCACGGGAGATGCTGTGTCCT 747
QY 130 GlyProProGlyLysGluGlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeu 149
DB 748 GGACAGCTTGGCTTAAAGGCTTCCCGAGACAGAGGCCCTCGCTGATTCAGAGAACG 807
QY 150 GlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGly 169
DB 808 CAAGGATG-----CCTGGATGCCAGGCTCAAGGCGCCAGGCGCCAGGCTCCAGGA 858
QY 170 ValProGlyMetProGlyProLysGlyProProGlyPro-----ProGlyProSer 186
DB 859 CTTCCGGCCAGCAGGCTGTATGGGCTCCAGGACTGCATGGATTCAGGAGTCCT 918
QY 187 GlyAlaValValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaProGluAsp 204
DB 919 GGCCCAAGAGGGGCTTGGGGCTGCCAGGAATCCCGAGCGCTGCAAGGCTGTGCTGTGAT 978
QY 205 AsnGlyCysPro 208
DB 979 AGAGGGGACCT 990
```

Search completed: March 21, 2003, 12:34:27

Job time : 46.3343 secs

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1934	99.7	2930	10	US-09-745-763-198	Sequence 1
2	474.5	24.5	1410	10	US-09-954-456-2274	Sequence 2
3	455	23.5	1133	9	US-09-924-340-57	Sequence 3
4	455	23.5	1133	9	US-09-992-600A-57	Sequence 4

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Alignment Scores:
Pred. No.:          Length:      2,49e-95      2930
Score:              1934.00      Matches:      341
Percent Similarity: 99.71%      Conservative: 0
Best Local Similarity: 99.71%      Mismatches: 1
Query Match:        99.69%      Indels:      0
DB:                  10         Gaps:         0

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-745-763-198 (1-2930)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1266 ATGCAACAAGATTGATGAGGTGAGGTAGACACTGAAGTAGCCCAACTATCAGTGATT 1325
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 1326 ATGGAAGAATGAGCTAGTAGCTCAAGTAGCTGACCTCATCAAGAATTTTCAATA 1385
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1386 CTACAAGGTCCACCGGGCCCCAGGGGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCT 1445
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 1446 GCGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGGGGGGCGCTGGACCACTGGCCCT 1505
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyCysArgGlyGly 100
Db 1506 GCGGGTGAGAGAGGCCCAATTGACCAAGCTGGTCCCGGAGAGCGTGGCGCAAGGA 1565
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1566 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAG 1625
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyGluGlyLeuProGlyPro 140
Db 1626 GCGCCCACTGGGACCCAGGCCCCCGGCCCAAGGAGAGAGGAGGAGTCCCGGCCCT 1685
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1686 CAGGGCCCCCTCGGCTTCAGGGACTTCAGGGACACCGTTGGGAGCGCTGGGTGCTGGA 1745
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1746 CCTCGGGGACTGCCAGCTTGCTGGGTACCAGGCATCCAGGCCCCCAAGGGCCCCCCC 1805
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1806 GGCCCTCTCTGGCCCATCAGGAGCGGTGGTGCCTCGCCCTGCAGAATCAGCAACCCCG 1865
QY 201 AlaProGluAspAsnGlyCysProProHisThrLysAsnPheThrAspLysCysThrTyr 220
Db 1866 GCACCGGAGGACAAATAGCTGCCCGCTCCTACTGGAAGAACTTCACAGACAAATGCTACTAT 1925
```

```

QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1926 TTTTCAGTTGAGAAACAAATTTTGGAGGTGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA 1985
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1986 CATCTTTTTCATAAACACACTAGAGAGAGACACGAATGGATAAAAAACAGATGGTAGG 2045
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2046 AGAGAGAGCCACTGGATCGCCTCACAGACTCAGAGCGGTGAAAATGAATGAAGTGCGTG 2105
QY 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2106 GATGGGACACTCTCCACACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTCAT 2165
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2166 GCCCATGGGCCAGGAGAAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGAAACGATTTC 2225
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2226 CAATGTGAAGACGTCAATAACTTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCT 2285
QY 341 AlaLeu 342
Db 2286 GCATTA 2291

RESULT 2
US-09-954-456-2274
; Sequence 2274, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2274
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-954-456-2274
```

```

Alignment Scores:
Pred. No.:          Length:      6.9e-18      1410
Score:              474.50      Matches:      111
Percent Similarity: 49.67%      Conservative: 41
Best Local Similarity: 36.27%      Mismatches:    132
```

```

; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
US-09-924-340-57

Alignment Scores:
Pred. No.: 6.26e-17 Length: 1133
Score: 455.00 Matches: 98
Percent Similarity: 47.46% Conservative: 14
Best Local Similarity: 41.53% Mismatches: 50
Query Match: 23.45% Indels: 74
DB: 9 Gaps: 7

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-924-340-57 (1-1133)

QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGly 1133
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140 CTCCCAGGCCCCCGACGCCAGCGGTCCTCCAGAGAGAGCGAGACCCCAAGT
                                     ||| ||||| |||||

QY 61 GlyProThrGlyAsnLysGlycLysGlyGluLysGlyGluProGlyProGlyPro
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 GGTGAGAAGGCGCACCAGGAGTTTCAAGGCCACGCCAGGCTTTCGCGGCCACCGG
                                     ||| ||||| |||||

QY 81 AlaglyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGly
   ||| ||| ||| ::||| ||||| ||||| ||||| |||||
Db 260 CCTGGATTCACGAGCAAGTTGGATCAGTGTGCCACCTGGCCCTCAAGCAGAG
                                     ||| ||||| |||||

QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGly
   ||||| ||| ::||| ||||| ||||| ||||| ||||| |||||
Db 320 AGCGAAGGGATTCCAGGGCCCATCAGGCTCGCTGGCTCCCTGGGCCACCGGGA
                                     ||| ||||| |||||

QY 120 -----GlnGlyProSer-----
   ||||| ||||| ::
Db 380 GGGATTACGGGCCCCCGCCGGTCTGGATGTTGGATGGGAGGATGGCAAGCTCT
                                     ||||| |||||

QY 124 ---GlyAspProGlyProProGlyProPro-----
   ||||| ||||| ||||| ||||| |||||
Db 440 AGGGGGGACCCCTGGTTCCTGCTGGCCGCCCTGGACTCATGGGACCACCGGGCTTT
                                     ||||| |||||

QY 133 -----GlyLysGluGlyLeuProGlyProGln-----
   ||| ||| ||||| ||||| |||||
Db 500 AAAACAGACACATCTCGGCTCTCCAGGACCTAAGGGTGACTGTGGCAACACCAAGT
                                     ||||| |||||

QY 145 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg
                                     ||||| |||||

```

RESULT 3
US-09-924-340-57
; Sequence 57, Application US/09924340
; Publication No. US20030027248A1

Db 560 GGCAGACTGGCGGCTGGCGCAGAGGTGAACCTGGTGCCATGGGACCCCGAGGAAGA 619
QY 165 ProGlyLeuPro----- 168
Db 620 CCCGGTCCCCGGGACAGCTTGGGCCACAGGGCCTCCAGGCCAGCCAGGACCGTGGG 679
QY 169 -----Gly 169
Db 680 ATCTCTGCAGTGGTCTGAAAGGAGACCGAGGAGCCACCGGAGAAAGGGGCTTCGAGGC 739
QY 170 ValProGlyMetProGlyProLysGlyProProGlyProPro-----GlyProSerGly 187
Db 740 CTCGCCAGGCGAGCCCGCCCGCCAGGTCTCAAGGTCTCTCCAGGCTATGGCAAGATGGGT 799
QY 188 AlaValValProLeuAlaLeuGlnAsnGluPro---ThrProAlaPro 202
Db 800 GCAACAGGACCAATGGCGCAGCAAGGCATCCCTGGGCATCCCTGGGGCCC 847

RESULT 4

US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephanie
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91 US4 DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 476..964
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
US-09-992-600A-57

Alignment Scores:
Pred. No.: 6,26e-17
Score: 455.00
Percent Similarity: 47.46%
Best Local Similarity: 41.53%
Query Match: 23.45%
DB: 9
Length: 1133
Matches: 98
Conservative: 14
Mismatch: 50
Indels: 74
Gaps: 7

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-992-600A-57 (1-1133)
QY 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 140 CTCGCCAGGCCCCCGCCAGCGGGTCCAGAGGAGAGCGAGGACCCCAAGGTAACCTCC 199
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro 80
Db 200 GGTGAGAAGGGCGACCGAGGATTTCAAGGCCACCCAGGCTTTCGGGCCCCACCGGTTCC 259
QY 81 AlaGlyGluArgGlyProLysGlyProAlaGlyProGlyGluArgGlyGlyGly 100
Db 260 CCTGGATTCCAGGCAAAAGTTGGATCACTGGCCCACTGGCCCTCAAGCAGAGAAGGCC 319
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro--- 119
Db 320 AGCGAAGGATTCGAGGCCCATCAGCCCTGCCTGGCTCCCTGGCCACCGGACCTCCT 379
QY 120 -----GlnGlyProSer----- 123
Db 380 GGGATTTCAGGGCCCCCGGCTCTGGATGGTTGGATGGGAAGGATGCCAAGCCTGGCTTG 439
QY 124 ---GlyAspProGlyProProGlyProPro----- 132
Db 440 AGGGGGACCTCTGCTCTGCTGGCCCCCTGGACTCATGGACCACCGGCTTTAAGGGG 499
QY 133 -----GlyLysGluGlyLeuProGlyProGln-----GlyProPro 144
Db 500 AAACAGGACATCTGCGCTCCCGAGGACCTAAGGTGACGTGGCAAAACAGGCTCTCCT 559
QY 145 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 164
Db 560 GGCAGCACTGGCGGCTGGCGCAGAGGTGAACCTGGTCCCATGGACCCAGGGAAGA 619
QY 165 ProGlyLeuPro----- 168
Db 620 CCCGGTCCCCCGGACACAGTTGGGCCACCGAGGCTCCAGGCCAGCCAGCAGCTGGG 679
QY 169 -----Gly 169
Db 680 ATCTCTGCAGTGGTCTGAAAGGAGAGCAGGAGCCAGCCAGGAGAAAGGGGCTTCGAGGC 739
QY 170 ValProGlyMetProGlyProLysGlyProProGlyProPro-----GlyProSerGly 187
Db 740 CTCGCCAGGCGAGCCCGCCCGCCAGGTCTCTCAAGGTCTCCAGGCTATGGCAAGATGGGT 799
QY 188 AlaValValProLeuAlaLeuGlnAsnGluPro---ThrProAlaPro 202
Db 800 GCAACAGGACCAATGGCGCAGCAAGGCATCCCTGGGCATCCCTGGGGCCC 847

RESULT 5

US-09-919-497-6
; Sequence 6, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 6158
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (2434)..(2434)
; OTHER INFORMATION: n = a, c, g or t/u
US-09-919-497-6

Pred. No.:	3.35e-16	Length:	6158
Score:	453.00	Matches:	130
Percent Similarity:	39.81%	Conservative:	38
Best Local Similarity:	30.81%	Mismatches:	76

Query Match:	23.35%	Indels:	180
DB:	10	Gaps:	18
US-09-763-712A-2_COPY_206_547 (1-342) x US-09-954-456-762 (1-6158)			
Qy 43	GlyProProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 62		
Db 4032	GGACCTCCAGGTCCCAAGGGCGCCAGTGATGATGCCCTTAAGGGTAACCCGGGTCCT 4091		
Qy 63	Thr-----		63
Db 4092	GTGGTTTTCCTGGAGATCCTGGTCCTCTGGGAACTTGCCCTGCAGTCAAGTGGT 4151		
Qy 64	-----GlyAsnLysGlnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAla 81		
Db 4152	GTGGTGGTGCACAAAGGTGAAGATGGAGATCCTGGTCAACGGGTCTCTCTGGCCCATCT 4211		
Qy 82	GlyGluArgGlyProIleGlyPro-----AlaGlyProProGlyGluArgGlyGly 98		
Db 4212	GGTGAGGCTGGCCACACAGGTCTCTGGAAACAGAGTCTCTGGAGTGCAGAGTGCA 4271		
Qy 99	LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro--- 117		
Db 4272	GAGGGAACAGCTCAAAAGGTCTAAGGGGAAGCAGGTGCAGAAAGTCTCTCTCGGA 4331		
Qy 118	-----GlyProGlnGlyProSerGlyAspProGlyPro----- 128		
Db 4332	AAAACCGGCCAGTCGGTCCTCAGGACCTGCAGAAAGCCTGTGCCAAGAGTCTTCGG 4391		
Qy 129	-----ProGlyProProGlyLysGluGlyLeu----- 137		
Db 4392	GGATCCCTGGTCTCTGGGGAACAAGTCTCCTCGAGTGCAGGCCAAGATGGACCA 4451		
Qy 138	ProGlyProGlnGlyProProGlyPhe----- 146		
Db 4452	CCTGGTCTATGGACCTCTCGCTTACCTGGTCTCAAGGTGACCCTGGCTCCAAAGGT 4511		
Qy 146	----- 146		
Db 4512	GAAGAAGGCATCCTGGTTTAATTGGCCTGATGTGCTCTCCAGGAGAACAAAGGGGAAAA 4571		
Qy 147	-----GlnGlyLeuGlnGlyThrValGlyGluPro-----GlyVal 158		
Db 4572	GGTGACGGAGGCTCCTCGAACTCAAGATCTCCAGAGCAAAAGGGGATGGGGAAAT 4631		
Qy 159	ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys--- 177		
Db 4632	CCTGGTCTCTGGTCCCTTAGGTCCACCTGGTCTCTCCAGGTTCACAGGTCTCTCAAGGC 4691		
Qy 177	----- 177		
Db 4692	CCAAGGGTAACAAAGGCTCTACTGGACCCCGTGGCCAGAAAGTGACAGTGGTCTTCCA 4751		
Qy 178	GlyProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGlu 197		
Db 4752	GGGCTCTCTGGGCTCCAGGTCCACTGGTGAATCATT-----CAG 4793		
Qy 198	ProThrPro-AlaProGluAspAsnGlyCysProProHistrLysAsnPhetheAspLys 217		
Db 4794	CCTTTACCAATCT-----TGTCCTCCAAAAACGA----- 4824		
Qy 217	sCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAs 237		
Db 4825	-----GAAGACATACTGAAGGCATGCAAG-- 4848		
Qy 237	pLysSerSerHisLeuValPheIleAsnThrArgGluGln----- 251		
Db 4849	-CAGATCCAGATGATAATATTCTGTACTCGGATGGAATGGAAGAAATATTGGTTCC 4907		
Qy 252	-----GlnTrpIleLysLysGlnMe 258		
Db 4908	CTCAATTCCCTGAACACAGACATCGAGATATGAATTTCCAATGGGTACTCAGACCAAT 4967		

Qy	258	tvalgYArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLy	278
:	:		
Db	4968	CCA-GCCCGAACTTGTAAAGAGCTGCAACTCAGCATCTGACTTCCACAGATGTGAATA	5026
Qy	278	sTrpLeuAsp-----GlyThrSerProAsp-----TyrLysAsnTrpLy	291
:	:	:	
Db	5027	TTGATTTGATCCTTAACCAAGTTGCTCAGGAGATTCCTTCAAAGTTTACTGTAATTTAC	5086
Qy	291	sAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGluAspCysAlaGlyLe	311
:	:		
Db	5087	ATCTGGT-----GGTGAGACTTGC-----	5105
Qy	311	uIleTyrAlaGlyClnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGl	331
:	:		
Db	5106	-ATTTATCCA-----GACAAATAATCTGAGGGAGTAAG-AATTTTCATCATGGC	5151
Qy	331	uLys 332	
:	:		
Db	5152	CAMA 5155	
RESULT 7			
US-09-954-456-782			
; Sequence 782, Application US/09954456			
; Patent No. US20020115057A1			
; GENERAL INFORMATION:			
; APPLICANT: Young, Paul			
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agen			
; TITLE OF INVENTION: Sets			
; FILE REFERENCE: 689290-76			
; CURRENT APPLICATION NUMBER: US/09/954,456			
; CURRENT FILING DATE: 2001-09-18			
; PRIOR APPLICATION NUMBER: US/60/233,617			
; PRIOR FILING DATE: 2000-09-18			
; PRIOR APPLICATION NUMBER: US/60/234,052			
; PRIOR FILING DATE: 2000-09-20			
; PRIOR APPLICATION NUMBER: US/60/234,923			
; PRIOR FILING DATE: 2000-09-25			
; PRIOR APPLICATION NUMBER: US/60/235,134			
; PRIOR FILING DATE: 2000-09-25			
; PRIOR APPLICATION NUMBER: US/60/235,637			
; PRIOR FILING DATE: 2000-09-26			
; PRIOR APPLICATION NUMBER: US/60/235,638			
; PRIOR FILING DATE: 2000-09-26			
; PRIOR APPLICATION NUMBER: US/60/235,711			
; PRIOR FILING DATE: 2000-09-27			
; PRIOR APPLICATION NUMBER: US/60/235,720			
; PRIOR FILING DATE: 2000-09-27			
; PRIOR APPLICATION NUMBER: US/60/235,840			
; PRIOR FILING DATE: 2000-09-27			
; PRIOR APPLICATION NUMBER: US/60/235,863			
; PRIOR FILING DATE: 2000-09-27			
; NUMBER OF SEQ ID NOS: 2276			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 782			
; LENGTH: 6728			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-954-456-782			
Alignment Scores:			
Pred. No.:		3,83e-16	Length: 6728
Score:		452.50	Matches: 99
Percent Similarity:		52.11%	Conservative: 12
Best Local Similarity:		46.48%	Mismatches: 53
Query Match:		23.32%	Indels: 49
DB:		10	Gaps: 8
US-09-763-7612A-2_COPY_206_547 (1-342) x US-09-954-456-782 (1-6728)			

QY 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspargGlySerGlnGlyProPro 60
:
:
:
Dd 1485 GTTCAAGACCCCTGGCCTGTGGAGAGGAAGAAAGCGAGGACTGCAGGTGAACC 1544


```

: PRIOR APPLICATION NUMBER: US/60/235,720
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,840
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,863
: PRIOR FILING DATE: 2000-09-27
: NUMBER OF SEQ ID NOS: 2276
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 786
: LENGTH: 5416
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-954-456-786

Alignment Scores:
Pred. No.:          9,05e-16          Length:          5416
Score:             444.00             Matches:         107
Percent Similarity: 47.15%             Conservative:    17
Best Local Similarity: 40.68%           Mismatches:     64
Query Match:       22.89%             Indels:         76
DB:                10                Gaps:          9

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-954-456-786 (1-5416)
QY 20 IleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThr 39
Db 2448 CTGCTCTCAGAGTGAATACCTGGCAGATGGTGGTG---CTCATG 2504
QY 40 IleLeuGlnGlyProGlyProArgGlyProArgGlyAspArg-----GlySer 56
Db 2505 GTGCTGTA-GGTGCCCTGGTCTGTCGAGCCACAGTGCAGCGGGGAGAGTGGGCT 2563
QY 57 GlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGly 76
Db 2564 GCTGCTCTGCTGCTGCTGCTGCTCGGGAAGCCCTGCTGGAACGTCGCGAGTGGT 2623
QY 77 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProPro-----93
Db 2624 CCTGCTGCCCCAACGGATTTGCTGGTCCGCTGCTGCTGCTGCTCAACCGGTCTAAA 2683
QY 94 GlyGluArgGlyLysGlySerLysGlySerGln-----105
Db 2684 GGAGAAGAGGAGGAAAGGCGCTAAGGCTGAACCGTGTGTTGTTGCCACAGGCCCC 2743
QY 106 -----GlyProLysGlySerArgGly 112
Db 2744 GTTGGAGTCTGTGGCCACGCTGCTCCAAATGTTCCCGCGTCTGCTGGAGTGGTGT 2803
QY 113 Ser-----ProGlyLysPro 117
Db 2804 GATGAGGCCCCCTGCTATGACTGTTCCCTGCTGCTGTCGACGAGTGGTCCCCCA 2863
QY 118 GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyLeu 137
Db 2864 GGACCTCTGTATTTCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2923
QY 138 ProGlyProGln-----Gly 142
Db 2924 CGTGGTCTCTGCTGTGACCAAGGTCAGTTGGCCGAACGTGGAGAAGTAGGTGCTGGT 2983
QY 143 ProGlyPheGlnGlyLeuGln-----GlyThrValGlyGluPro 156
Db 2984 CCCCCTGGTCTGCTGTGAGAAGGTCCTCTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCT 3043
QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro 176
Db 3044 GGCACTCCAGTCTCAGGGTCTTCTGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3103
QY 177 LysGlyProProGlyProProGlyProSerGlyAlaVal-----ValProLeuAla 193
Db 3104 AGAGGTGAACGTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3163
QY 194 LeuGlnAsnGluPro-----ThrProAlaProGluAspAsn 205

```

Db 2864 GGACCTCTGGTATTTCTGGCCCTCCTGGTCCCTGGTCTGGGAAAGAGGCGTT 2923
QY 138 ProGlyProGln-----Gly 142
Db 2924 CGTGGTCTCGTGGTACCAAGGTCCAGTGGCCGAACTGGAGAGTAGTGCAGTTGGT 2983
QY 143 ProGlyPheGlnGlyLeuGln-----GlyThrValGlyGluPro 156
Db 2984 CCCCCTGGCTTGGCTGGTGAAGAGGTCCCTCTGGAGAGGCTGGTACTGCTGGACCTCCT 3043
QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro 176
Db 3044 GGACTCCAGGCTCAGGCTCTCTGGTGGTCTCTGGTATTCTGGGTCTCCTGGCTCG 3103
QY 177 LysGlyProGlyProGlyProGlyProSerGlyAlaVal-----ValProLeuAla 193
Db 3104 AGAGGTGAACGTGGTCTACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3163
QY 194 LeuGlnAsnGluPro-----ThrProAlaProGluAspAsn 205
Db 3164 ATTGGCGGCCCTCCTGGGCGCGTGGTCTCCTGGTGGTGGTGGTGGTGGTGGTGGTGG 3223
QY 206 GlyCysPro 208
Db 3224 GGTGCTCCT 3232

RESULT 11
US-10-044-090-104
; Sequence 104, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 104
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.19
; LOCATION: 1492-1495, 1663, 1669, 4012, 4055
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-104

Alignment Scores:
Pred. No.: 1,11e-15 Length: 4149
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 22.71% Indels: 53
DB: 12 Gaps: 6
US-09-763-712a-2_COPY_206_547 (1-342) x US-10-044-090-104 (1-4149)
QY 43 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
Db 1307 GGTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1366
QY 63 Thr-----Gly 64
Db 1367 AATGGAATTTCTGGTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1426
QY 65 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyProGlyPro 84
Db 1427 GCCAAGGGCCCTAAGGGTGAAGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1486

QY 85 GlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGlySerLysGlySer 104
Db 1487 GGCCCNNGGTCCAAATGGTCCCGGTCCTGCTGGAAGTCTGGTATGAGGCCCC 1546
QY 105 GlnGlyProLysGlySerArgGlyProGlyLysProGlyProGlnGlyPro----- 122
Db 1547 CTGGTATGACTGGTTTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1606
QY 123 ---SerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGln 141
Db 1607 ATTCTGGCCCTCCTGGTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1666
QY 142 -----GlyProProGlyPhe 146
Db 1667 GGNAGCAAGGACAGCAGCGCCGACCTGGAGAAAGTAGGAGCACCAGGCTCCCGCTGGCTTC 1726
QY 147 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 160
Db 1727 GCTGGTGAAGAGGTCCTCTGGAGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1786
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1787 CCTCAGGCTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1846
QY 181 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 197
Db 1847 GGTCTACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1906
QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
Db 1907 CCTGGGCGCGTGGTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1963

RESULT 12
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Vockley, Darci T.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: unsure
; LOCATION: (1)...(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Alignment Scores:
Pred. No.: 1,32e-15 Length: 5086
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 22.71% Indels: 53
DB: 10 Gaps: 6
US-09-763-712a-2_COPY_206_547 (1-342) x US-09-880-107-3947 (1-5086)
QY 43 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62

```
Db 2246 GGTCTGCTGGTCTCGGGAAGCCCTGGTGAACGTGGCGAGGTGCTCTGCTGCGCC 2305
QY 63 Thr-----Gly 64
Db 2306 AACGGATTTGCTGGTCCGCTGGTCTGCTGCTCAACCGGGTCTAAAGGAGAAAGAGA 2365
QY 65 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyGluArg 84
Db 2366 GCCAAAGGGCCTAAGGCTGAACCGTGTGTGTGCTCCACAGGCCCTGTGAGAGTGTCT 2425
QY 85 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysSerLysGlySer 104
Db 2426 GGCCCCAGCTGGTCCCAATAGTCCCGCTGCTGCTGGAAGTGTGATGAGGCCCC 2485
QY 105 GlnGlyProLysSerArgGlySerProGlyLysProGlyProGlyProGlyProGlyPro 122
Db 2486 CCTGGTATGACTGTTTCCCTGGTGTGCTGAGCGACTGTGCTCCACAGGACCTGTGT 2545
QY 123 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 141
Db 2546 ATTCTGGCCCTCTGCTGCTCCCTGCTGCTGCTGGAAGAGGCTTCTGCTGCTCTCT 2605
QY 142 -----GlyProProGlyPhe 146
Db 2606 GTGACCAAGGTCAGTTGGCCGAAGTGAAGAGTAGTGGTGGTCCCTGCTGCTCTC 2665
QY 147 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 160
Db 2666 GCTGGTGAGAGGGTCCCTCTGGAGAGGTGTGTGCTGAGCTCTGCTGCTGCTGCT 2725
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 2726 CCTCAGGGTCTTCTTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2785
QY 181 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 197
Db 2786 GGTCTACCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2845
QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
Db 2846 CTGGGGCCCGTGGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCT 2902

RESULT 13
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
```

```
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.: 1.33e-15 Length: 5145
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 22.71% Indels: 53
DB: 9 Gaps: 6

US-09-763-712a-2_copy_206_547 (1-342) x US-09-925-299-206 (1-5145)
QY 43 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
Db 2248 GGTCTGCTGGTCTCGGGAAGCCCTGGTGAACGTGGTGAAGTGGTCTGCTGCTGCT 2307
QY 63 Thr-----Gly 64
Db 2308 AATGGATTTGCTGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2367
QY 65 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyProGlyGluArg 84
Db 2368 GCCAAAGGGCCTAAGGCTGAACCGTGTGTGTGCTCCACAGGCCCTGTGAGAGTGTCT 2427
QY 85 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysSerLysGlySer 104
Db 2428 GCCCCAGCTGGTCCCAATAGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2487
QY 105 GlnGlyProLysSerArgGlySerProGlyLysProGlyProGlyProGlyProGlyPro 122
Db 2488 CCTGGTATGACTGTTTCCCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2547
QY 123 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 141
Db 2548 ATTCTGGCCCTCTGCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCT 2607
QY 142 -----GlyProProGlyPhe 146
Db 2608 GTGACCAAGGTCAGTTGGCCGAAGTGAAGTAGTGGTGGTCCCTGCTGCTGCTGCTGCT 2667
QY 147 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 160
Db 2668 GCTGGTGAGAGGGTCCCTCTGGAGAGGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2727
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 2728 CCTCAGGGTCTTCTTGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2787
QY 181 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 197
Db 2788 GGTCTACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2847
QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
Db 2848 CTGGGGCCCGTGGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCT 2904

RESULT 14
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```

```

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 5432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
; -10-044-090-22
;

```

Alignment Scores:		
Pred. No.:	1.39e-15	5432
Score:	440.50	95
Percent Similarity:	49.32%	Conservative: 13
Best Local Similarity:	43.38%	Mismatches: 58
Query Match:	22.71%	Indels: 53
DB:	12	Gaps: 6

US-09-763-712A-2_COPY_206_547 (1-342) x US-10-044-090-22 (1-5432)

```

US-09-763-712A-2_COPY_206_547 (1-342) x US-10-044-090-22 (1-5432)

Qy 43 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 62
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2579 GGTCTGCTGGTCTCTGGGGAAGCCCTGCTGAACGTGGTGGAGTGGTCTGCTGGCCCC 2638
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 63 Thr-----Gly 64
      |||||

Db 2639 AATGGATTTGCTGGTCTGGTGTCTGCTCAACCTGGTGTCAAAGGAGAAAGAGA 2698
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 65 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArg 84
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2699 GCCAAGGGGCTAAGGCTGAAAACGGTGTGTTGGTCCACAGGCCCTGTTGGAGCTGCT 2758
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 85 GlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySer 104
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2759 GGCCCAAGTGGTCCAAATGGTCCCCCGCTCTGCTGGAAGCTGGTGGATGGAGGCCCC 2818
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 105 GlnGlyProLysGlySerArgLysProGlyLysProGlyProGlnGlyPro----- 122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2819 CTTGTATGACTGGTTTCCCTGGTGTCTGGTGGACGGACTGGTCCCCCAGGACCTCTGGT 2878
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 123 ---SerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGln 141
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2879 ATTCTGGCCCTCTGCTGCCCTGCTCTGCTGGAAAGAGGGCTTCGTGGTCTCTGCT 2938
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 142 -----GlyProProGlyPhe 146
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2939 GGTGACCAAGGTCCAGTTGGCCGAAGTGGAGAAGTAGGTGCAGTTGGTCCCCCTGGCTTC 2998
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 147 GlnGlyLeuGln-----GlyThrValGlyLupProGlyValProGly 160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2999 GCTGGTGAGAAGGGTCCCTCTGGAGAGGCTGGTACTGCTGGACCTCTGGCAGCTCCAGGT 3058
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3059 CCTCAGGTCCTTGGTGTCTCTGGTATTCCTGGGTCTCCCTGGCTCGAGAGGTGAAGCT 3118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 181 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 197
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3119 GGTCTACCAAGTGTGCTGGTGTGCTGGGTGAACCTGGTCTCTTTGGCATTTCCCGGCCCT 3178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Qy 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
|||
|||
:::|
|||

QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
 ||| ::||| ||||| |||
Db 3179 CCTGGGGCCCGTGGTCTCTGGTGCTGTGGGTAGTCTCGAGTCAACGGTGCTCCT 3235

Search completed: March 21, 2003, 12:32:28
Job time : 186.172 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:08:41 ; Search time 1130.43 Seconds
(without alignments)
4899.781 Million cell updates/sec

Title: US-09-763-712A-2_COPY_206_547
Perfect score: 1940
Sequence: 1 MQQDLMSRLDTEVANLSVI.....EDVNNFICEKDRVTLSSAL 342

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09763712/runat_14032003_100950_18109/app_query.fasta_1.1877
-DB=EST -QFMT=fastap -SURFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -CGN_1_1_4575 -runat_14032003_100950_18109 -NCPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1801	92.8	3305	11	BC009162	BC009162 Mus muscu
2	1256	64.7	861	14	BQ713873	BQ713873 AGENCOURT
3	1061	54.7	906	13	BI456109	BI456109 603172765
4	1022	52.7	580	12	BE910803	BE910803 601661855
5	1007	51.9	668	14	BQ674807	BQ674807 AGENCOURT
6	957	49.3	601	10	BE290299	BE290299 601089246
7	937	48.3	957	14	BQ955927	BQ955927 AGENCOURT
8	916.5	47.2	967	9	AL568743	AL568743 AGENCOURT
9	842	43.4	552	10	AW958053	AW958053 EST370123
10	831	42.8	638	10	BB248064	BB248064 BB248064
11	767	39.5	808	14	BQ71366	BQ71366 UI-M-F10-
12	696.5	35.9	936	14	BQ891432	BQ891432 AGENCOURT
13	669	34.5	500	14	BQ676508	BQ676508 UI-E-E30-
14	663	34.2	507	14	BQ713891	BQ713891 UI-E-E30-
15	627.5	32.3	591	13	BI445884	BI445884 dal33d10.
16	600	30.9	357	10	AW379436	AW379436 CM4-HT024
17	594	30.6	350	9	AA012704	AA012704 RPUI402CG
18	588	30.3	339	9	AA304251	AA304251 EST17006
19	560	28.9	517	9	AL543000	AL543000 AL543000
20	537	27.7	310	9	AA361740	AA361740 EST71069
21	478	24.6	943	14	BQ887163	BQ887163 AGENCOURT
22	460	23.7	441	10	BB749743	BB749743 BB749743
23	459.5	23.7	1082	14	BM907108	BM907108 AGENCOURT
24	458.5	23.6	569	10	AV616076	AV616076 AV616076
25	456.5	23.5	792	13	BI456015	BI456015 603170460
26	456	23.5	420	12	BE929813	BE929813 RC3-GN004
27	453.5	23.4	896	14	BQ955616	BQ955616 AGENCOURT
28	453.5	23.4	877	12	BQ681943	BQ681943 602629884
29	452	23.3	688	13	BI249235	BI249235 602995511
30	452	23.3	3951	11	BC013626	BC013626 Mus muscu
31	451.5	23.3	921	14	BQ921102	BQ921102 AGENCOURT
32	451.5	23.3	1192	14	BM905999	BM905999 AGENCOURT
33	449.5	23.2	907	14	BQ222937	BQ222937 AGENCOURT
34	448	23.1	946	14	BQ942234	BQ942234 AGENCOURT
35	447.5	23.1	1029	17	CNS04VYG	AL309697 Tetraodon
36	446.5	23.0	608	13	BM425818	BM425818 pbf2c.pk0
37	446	23.0	651	10	BB613976	BB613976 BB613976
38	446	23.0	819	12	BQ679800	BQ679800 602626481
39	446	23.0	1120	14	BM809933	BM809933 AGENCOURT
40	445.5	23.0	1014	14	BQ942135	BQ942135 AGENCOURT
41	445	22.9	972	14	BQ896478	BQ896478 AGENCOURT
42	444.5	22.9	678	13	BM486385	BM486385 pgmlc.pk0
43	444.5	22.9	781	13	BI454661	BI454661 603170558
44	444	22.9	429	12	BE929798	BE929798 RC3-GN004
45	444	22.9	991	14	BQ713634	BQ713634 AGENCOURT

ALIGNMENTS

RESULT 1
BC009162
LOCUS BC009162 3305 bp mRNA linear HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to scavenger receptor with C-type lectin,
clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 5 Row: h Column: 10

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis

This clone has the following problem: incomplete processing.

FEATURES

source
Location/Qualifiers

```

1..3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCL CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 1031 a 757 c 760 g 757 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 2,02e-108 Length: 3305
Score: 1801.00 Matches: 313
Percent Similarity: 95.91% Conservative: 15
Best Local Similarity: 91.52% Mismatches: 14
Query Match: 92.84% Indels: 0
DB: 11 Gaps: 0

```

US-09-763-712A-2_COPY_206_547 (1-342) x BC009162 (1-3305)

```

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1296 ATGCAGCAAGACATGATGAGGTCAAGTATAGACACTGAAGTGGCCCACTATCAGTGGTT 1355

QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1356 ATGGAAGAGATGAACCTGGTGTACTCCAAAGCAGCGTCAGTCATCAAGAACTTTACCAAT 1415

QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyValAspArgSerGlnGlyProPro 60
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1416 CTACAAGTCTCTCTGGCCCCAGAGGTCCAAAGGTGACAGAGATCTCAGGAGCCACCT 1475

QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1476 GGTCCAACGTGGCAACAAGGACAGAAAGGAGAGAGAGAGCGCTGGTCCACCTGGCCCT 1535

QY 81 AlaGlyGluArgGlyProLeuGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1536 GCGGGGTGAGAGGGGCCAATTTGGACCAAGTCCCTCGGTGGAGAGCGTGGCAGCAAGGA 1595

QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro 120
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1596 TCCAAAGGCTCAGAGGTCCTCCAAAGGATCTGTGGGTCTCCCGAGGAAGCTGGCCCTCAA 1655

```

```

QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 140
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1656 GGACCTAGTGGGAGCCAGGACCCAGGAGTCCACAGGCAAGGATGAGCTCCCTGGCCCT 1715

QY 141 GlnGlyProProGlyPheGlnGlnGlyLeuGlnGlnGlyThrValGlyGluProGlyValProGly 160
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1716 CAGGGCCCTCTGGCTTCCAGGGACTACAGGGCAGCTGTGGGTGAGCTGGAGTACCTGGA 1775

QY 161 ProArgGlyLeuProGlyProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1776 CCTGGGGGTGGCAGGCTTGCAGGGGTGCCAGGAGTGCCTGGGCTTAAGGACCACCT 1835

QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1836 GGCCCTCCAGGCCCCCTCAGGAGCAATGGAGCCATTGGCTCGCAGATGAACCAACCCCA 1895

QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTyrTyr 220
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1896 GCATCAGAGGTCAACGGATGTCCACCTCAGTGAAGAACTTCACAGATAAATGCTACTAT 1955

QY 221 PheSerValGluLysGluIlePheGluAspAlaLysIlePheCysGluAspLysSerSer 240
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 1956 TTTTCATTGGAAGAAAGAAATTTTGAAGATGCTAAAGCTTTTCTGTGAAGACAAATCTTC 2015

QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2016 CATCTCGTTTCATNAACTCAAGAGAAGAACAGCAATGGATAAAGCATACCGTGGG 2075

QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2076 AGAGAAAGCCATTGGATCGCCCTCACAGACTCAGAAAGCAAGAAAGCAAGTGGAGTGGCTA 2135

QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2136 GACGGGTACCTGTTGATTACAAAAGCTGGAAGCTGGCAACCACTAACTACTGGGCAGT 2195

QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2196 GGCCATGGCCAGGAGAACAGACTGTGCTGGCTTATACGAGGACAGTGGATGACTTC 2255

QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 2256 CAGTGTGATGAATCAATAACTTTCATTGTGAGAAGAAAGGAGGAGGAGTACCATCATCC 2315

QY 341 AlaLeu 342
|||
Db 2316 ATATTA 2321

```

RESULT 2

B0713873

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B0713873
AGENCOURT_8482828 NIH_MGC_129 Mus musculus cdna clone IMAGE:6306240
5', mRNA sequence.

B0713873.1 GI:21852772

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 861)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

cDNA Library Preparation: ResGen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:


```

http://image.llnl.gov
Plate: L1AM13721 row: n column: 01
High quality sequence stop: 598.
FEATURES
    source
        1. 861
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:6306240"
            /lab_host="NIH_MGC_129"
            /note="Organ: olfactory epithelium; Vector:
            pCMV-SPORT6.1 ccdB; Site_1: EcoRV; Site_2: NotI; Cloned
            unidirectionally. Primer: Oligo dt. Average insert size
            2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
            is a NIH_MGC Library."
BASE COUNT    221 a 253 c 239 g 147 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      3.18e-73      Length:      861
Score:          1256.00      Matches:      229
Percent Similarity: 91.01%      Conservative: 14
Best Local Similarity: 85.77%      Mismatches: 20
Query Match:      64.74%      Indels:      4
DB:              14      Gaps:      0

US-09-763-712a-2_COPY_206_547 (1-342) x BQ713873 (1-861)
QY    5  LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMet 24
Db    13 ATGATGAGTCAAGGTAGACACTGAAGTGGCCAACTTATCATGTTGTTGAAGAGATG 72
QY    25 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro 44
Db    73 AAACGTGTTGACTCCAAGCAGCGTCAGCTCATCAAGAAGCTTTACCATTCTCAAGGTCT 132
QY    45 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGly 64
Db    133 CTGCGCCCGAGAGGTCCAAAGAGGTACAGAGATCTCAGGGACCACTGTGTCCAACTGGC 192
QY    65 AsnLysGlyGlnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAlaGlyGluArg 84
Db    193 AACAGGGACAGAGGAGAGGAGGAGGAGCTGTGTCACCTGGCCCTCGGGTGAAGG 252
QY    85 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer 104
Db    253 GCCAATTTGGACCACTCGGCCCTCTCTGGAGAGCGTGGCAGCAAGGATCCAAAGGCTCA 312
QY    105 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGly 124
Db    313 CAGGGTCCCAAGGATCTGTGGGTCCCGAGGAAAGCCTGGCCCTCAAGGACCTAGTGGG 372
QY    125 AspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProPro 144
Db    373 GACCAGGACCACTGAGTCCAGCAGGAGGATGACTCCCTGGCCCTCAGGGCCCTCTCT 432
QY    145 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 164
Db    433 GCCTTCCAGGACTACAGGCACTGTGGTGAGCTGGAGTACCTGGACCTCGGGGTG 492
QY    165 ProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGly 184
Db    493 CCAGGCTTCCAGGGGTGCGAGGCATGCTGGGGTCCCGAGGAAAGCCTGGCCCTCCAGGC 552
QY    185 ProSerGlyAlaValAlaProLeuAlaLeuGlnAsnGluProThrProAlaProGluAsp 204
Db    553 CCTCAGGAGCAATGAGCCATTGTGCTCTGCAAAATGAACCAACCCANCATCAGAGTTC 612
QY    205 AsnGlyCysProHisHisLysAsnPheThrAspLysCysTyrTyrPheSerVal-Gl 224
Db    613 AACGATGTCCGCCCTCAGTGGAGAACTTCACAGATAAATGCTACTATTTTCTTGGCA 672
QY    224 ulysGluIlePheGluAspAlaLysLeu-PheCysGluAspLysSerSerHisLeuValP 244

```

```

Db    673 AAAAAAATTTTCGAAATGCTACCTTTTCTCGGAACAAAAATCTCCCATCCGGGT 732
QY    244 heileAsnThrArgGluGlnGlnTrp-Ile-LysLysGlnMetValGlyArgGluSe 263
Db    733 TCATAACCCCAAGAAAAACAGCCATGGATAAAAACCTACCTTCGGGGACCAAAAC 792
QY    263 rHisTrpIleGly 267
Db    793 CGGCTGGATCGGG 805

RESULT 3
BI456109 603172765F1 NCI_CGAP_Mam5 Mus musculus cDNA linear EST 21-AUG-2001
LOCUS    mRNA sequence.
DEFINITION
ACCESSION BI456109
VERSION   BI456109.1 GI:15246765
KEYWORDS EST.
SOURCE    house mouse.
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
  1 (bases 1 to 906)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL Unpublished (1999)
  COMMENT Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov
          Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: L1AM1636 row: b column: 17
          High quality sequence stop: 730.
FEATURES
    source
        1. 906
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:5251888"
            /clone_lib="NCI_CGAP_Mam5"
            /tissue_type="tumor, gross tissue"
            /dev_stage="7 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies. Investigators
            providing samples: Lothar Hennighausen/Robin Humphreys,
            NIH"
BASE COUNT    247 a 228 c 256 g 175 t
ORIGIN
Alignment Scores:
Pred. No.:      1.97e-60      Length:      906
Score:          1061.00      Matches:      203
Percent Similarity: 87.76%      Conservative: 12
Best Local Similarity: 82.86%      Mismatches: 26
Query Match:      54.69%      Indels:      6
DB:              13      Gaps:      2

US-09-763-712a-2_COPY_206_547 (1-342) x BI456109 (1-906)
QY    1  MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db    123 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTGGCCCACTTATCAGTGT 182
QY    21  MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db    183 ATGGAAGAGATGAAACTGTTGACTCCAAAGCAGCGTCACTCATCAAGAACTTTACCAT 242

```


KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 668)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LUCM2377 row: c column: 09 High quality sequence stop: 656. Location/Qualifiers 1. .668 /organism="Homo sapiens" /db xref="taxon:9606" /clone="IMAGE:6212672" /clone_lib="NIH_MGC_102" /tissue_type="epidermoid carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library." 216 a 148 c 167 g 137 t
FEATURES	
source	
BASE COUNT	
ORIGIN	

Alignment Scores:			
Pred. No.:	5,21e-57	Length:	668
Score:	1007.00	Matches:	174
Percent Similarity:	99.43%	Conservative:	0
Best Local Similarity:	99.43%	Mismatches:	1
Query Match:	51.91%	Indels:	0
DB:	14	Gaps:	0
US-09-763-712A-2_COPY_206_547 (1-342) x BQ674807 (1-668)			
QY	168	ProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyProSerGly	187
Db	3	CCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCGGGCCCTCTTGGCCCATCAGGA	62
QY	188	AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCys	207
Db	63	CGGTGTGTGCCCTTGGCCCTGCAAGATGAGCCAAACCCGCGACCGGAGGACAAATAGCTGC	122
QY	208	ProProHistrpLysAsnPheThrAspLysCysTyrrTyrrPheSerValGluLysGluIle	227
Db	123	CCGCCTCAGCTGGAAGAACTTTCACAGACAAATGCTACTATTTTTTCACTTTCAGAAAAGAAAT	182
QY	228	PheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThr	247
Db	183	TTTGAGGATGCAAGCTTTTCTGTGAAGACAAGCTTTCATCTTCTGTTTCATAAACACT	242
QY	248	ArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGly	267
Db	243	AGAGAGGAACAGCAATGGATAAAAAACAGATGATGGAGAGAGAGCCACTGGATCGCG	302
QY	268	LeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyr	287
Db	303	CTCACAGACTCAGAGGCTGAAATGAATGAATGAAGTGGCTGGATGGGACATCTCCAGACTAC	362

Qy	288	LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAasp	307
Db	363	AAAAATTGGAAAGCTGGACACCGCATTAATTCGGGTCAATGCCTATGGCCATGGGCCAGAGAAGAC	422
Qy	308	CysAlaGlyLeuIleTyTAlacGlyClnTrpAsnAspPheGlnCysGluAaspValAsnAsn	327
Db	423	TGTGCTGGGTGAATTATGCTGGGCAGTGGAACAATTTCCAATGTGAAGACGTCAATAAAC	482
Qy	328	PheIleCysGluLysASPAArgGLUThrVallLeuSerSeraLaLeu	342
Db	483	TTCATTTCGAAAAGACAGCAGGACACTGTCATCTGCATTA	527
<hr/>			
RESULT 6			
LOCUS BE290299 601 bp mRNA linear EST 26-OCT-2000			
DEFINITION 601089246f1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5',			
mRNA sequence.			
ACCSSION BE290299			
VERSION BE290299.1 GI:9171250			
KEYWORDS EST.			
SOURCE house mouse.			
ORGANISM Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
1 (bases 1 to 601)			
NIH-MGC http://mgc.nci.nih.gov/.			
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished (1999)			
CONTACT: Robert Strausberg, Ph.D.			
Email: rcgaps@remail.nih.gov			
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys			
CDNA Library Preparation: Life Technologies, Inc.			
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Incyte Genomics, Inc.			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: LLAM8516 row: k column: 12			
High quality sequence stop: 551.			
<hr/>			
FEATURES			
source Location/Qualifiers			
1..601			
/organism="Mus musculus"			
/strain="C57BL/6J"			
/db_xref="taxon:10090"			
/clone_image="IMAGE:3484163"			
/clone_lib="NCI_CGAP_Mam5"			
/tissue_type="tumor, gross tissue"			
/dev_stage="7 months"			
/lab_host="DH10B"			
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;			
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt."			
Library constructed by Life Technologies. Investigators			
providing samples: Lothar Hennighausen/Robin Humphreys,			
NIH"			
<hr/>			
BASE COUNT	156 a	161 c	179 g 105 t
ORIGIN			
<hr/>			
Alignment Scores:			
Pred. No.:	8.97e-54	Length:	601
Score:	957.00	Matches:	183
Percent Similarity:	93.10%	Conservative:	6
Best Local Similarity:	90.15%	Mismatches:	11
Query Match:	49.33%	Indels:	6
DB:	10	Gaps:	0
<hr/>			
US-09-763-712A-2_COPY_206_547 (1-342) x BE290299 (1-601)			
<hr/>			
Qy	64	GlyAsnLysGly-GlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGl	83
Db	2	GGCACRACAGGTACAGAAGGAGAGAGCGCTGTCCTGCCCTCGGGTGA	61
Qy	83	uARGglyProileGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysgl	103
Db			

REFERENCE 1 (bases 1 to 967)
 Li.W.B., Gruber.C., Jesse.J. and Polayes.D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 JOURNAL
 COMMENT
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 Location/Qualifiers
 source
 1..967
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DE005YH04"
 /clone_lib="LTI_FL002_PL1"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
 ORIGIN

Alignment Scores:
 Pred. No.: 6.06e-51 Length: 967
 Score: 916.50 Matches: 174
 Percent Similarity: 91.15% Conservative: 1
 Best Local Similarity: 90.62% Mismatches: 14
 Query Match: 47.24% Indels: 5
 DB: 9 Gaps: 3

US-09-763-712A-2_COPY_206_547 (1-342) x AL568743 (1-967)

QY 151 GlyThrValGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyVal 170
 |||||
 Db 947 GCACCGTGGAGCGCTGGTCTCT---GGACCTCGGAGTGCAG---CTTGCTGGGTA 894
 |||||

QY 171 ProGlyMetProGlyProGlyProGlyProGlyProGlyProSerGlyAlaValVal 190
 |||||
 Db 893 CCAGGCATGCCAGSCCAAGG---CCCCCGCCCTCTCTGC-CCATCAGGACGGGTG 838
 |||||

QY 191 ProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHis 210
 |||||
 Db 837 CCCCTGSCCTGCAGATGAGCAACCCCGGACCGGAGCAATGGCTGCCGCTCAC 778
 |||||

QY 211 TrpLysAsnPhThrAspLysCysTyrTyrPheSerValGluLysGluPheGluAsp 230
 |||||
 Db 777 TGAAGAAGCTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTT 718
 |||||

QY 231 AlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGlu 250
 |||||
 Db 717 GCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTTATAAAGCTAGAG 658
 |||||

QY 251 GlnGlnTrpIleLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAsp 270
 |||||
 Db 657 CAGCAATGATAAAGAAACAGATGTAGGAGAGAGACCTGGATCGGCTCACAGAC 598
 |||||

QY 271 SerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrp 290
 |||||
 Db 597 TCAGAGCGTGAATGAATGAAGTGGTGGATGGGACATCTCCAGACTACAAAATGG 538
 |||||

QY 291 LysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 310
 |||||
 Db 537 AAAGCTGGACAGCCGGAATAACTGGGTGTCATGC-CATGGCCAGGAGAACTGTGCTGG 479
 |||||

QY 311 LeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhelLecys 330
 |||||
 Db 478 TTGATTTATGCTGGCAGTGGGAACGATTTCCAATGTGAAGACGTCAATAACTTCATTG 419
 |||||

QY 331 GluLysAspArgGluThrValLeuSerSerAlaLeu 342
 |||||

Db 418 GAAAGACAGAGGAGACAGYACTGTCTCATCTGCATTA 383
 |||||

RESULT 9
 AW958053 552 bp mRNA linear EST 01-JUN-2000
 LOCUS EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
 DEFINITION AW958053
 ACCESSION AW958053
 VERSION AW958053.1 GI:8147736
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
 ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
 Quackenbush,J.
 TITLE Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 JOURNAL Unpublished (2000)
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 115
 Seq primer: Reverse.
 Location/Qualifiers
 source
 1..552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGE"
 /note="Vector: pBluescriptSKm"

BASE COUNT 176 a 112 c 143 g 121 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2.84e-46 Length: 552
 Score: 842.00 Matches: 149
 Percent Similarity: 98.68% Conservative: 1
 Best Local Similarity: 98.03% Mismatches: 1
 Query Match: 43.40% Indels: 1
 DB: 10 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AW958053 (1-552)

QY 192 LeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrp 211
 |||||

Db 1 CTGGCCCTGCAGAAAGCTGCTACTATTTTTCAGTTGAGAAAGAAATTTT 60
 |||||

QY 212 LysAsnPhThrAspLysCysTyrTyrPheSerValGluLysGluPheGluAspAla 231
 |||||

Db 61 AAGAAGCTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTT 120
 |||||

QY 232 LysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGln 251
 |||||

Db 121 AAGCTTTTCTGTGAGACAAAGTCTTCACATCTTGTGTTTATAAAGCTAGAGAG 180
 |||||

QY 252 GlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSer 271
 |||||

Db 181 CAATGGATAAAGAAACAGATGTAGGAGAGAGACCTGGATCGGCTCACAGACTCA 240
 |||||

QY 272 GluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLys 291
 |||||

Db 241 GAGCGTGAATGAATGAAGTGGTGGATGGACATCTCCAGACTACAAAATTTGAAA 300
 |||||

QY 292 AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 311
 |||||

Db 301 GCTGGACAGCCGGAATAACTGGGTGTCATGCCCATGGCCAGGAGAACTGTGCTGGGTTG 360
 |||||

[illegible]

```

RESULT 11
BQ771366      808 bp      mRNA      linear      EST 26-JUL-2002
LOCUS        UI-M-F10-byu-g-09-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
DEFINITION   IMAGE:5702432 5', mRNA sequence.
ACCESSION   BQ771366
VERSION     BQ771366.1 GI:21979842
KEYWORDS     EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgi.nci.nih.gov/.
            1 (bases 1 to 808)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
Seq primer: pYX-5.
            Location/Qualifiers
            1..808
            /organism="Mus musculus"
            /strain="CS7BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:5702432"
            /clone_lib="NIH_BMAP_F10"
            /tissue_type="whole brain"
            /dev_stages="embryo 12.5dpc"
            /note="Organ: Brain; Vector: pYX-Asc; Site:1: Ecor I;
            Site:2: Not I; The library was constructed according
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with Ecor I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is CAGCCAGCAC. This library was created for the University
            Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
            program coordinator."
BASE COUNT   270 a 158 c 173 g 206 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      3,22e-41      Length:      808
Score:          767.00      Matches:    131
Percent Similarity: 93.3%      Conservative: 9
Best Local Similarity: 87.3%      Mismatches: 10
Query Match:    39.54%      Indels:    0
DB:            14          Gaps:      0

US-09-763-712A-2_COPY_206_547 (1-342) x BQ771366 (1-808)
QY 193 AlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLys 212
      |||||||
Db 2 GCTCTGCAGATGACCAACCCAGCATCAGAGTGCACGGATGTCGCGCTCAGTGGAG 61

QY 213 AsnPheThrAspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLys 232
      |||||||
Db 62 AACTTTCACAGATAAATGCTACTATTATTTTCATTTGGAAGAAATTTTGAAGATGCTAAG 121

```

```

QY 233 LeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGln 252
      |||||||
Db 122 CTTTCTCTGTGAACAATCTTCCCATCTCGTTTTCATAAATCAAGAGAGAACAGCAA 181
      |||||||
QY 253 TrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGlu 272
      |||||||
Db 182 TGGATAAAAAAGCATACCGTGGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGAA 241
      |||||||
QY 273 ArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAla 292
      |||||||
Db 242 CAGGAAGCGAATGAAGTGGCTAGACGGGTCCCTGTTGATTACAAAACCTGGAAGCT 301
      |||||||
QY 293 GlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIle 312
      |||||||
Db 302 GGACAACCCAGATAACTGGGCAGTGGCCATGGCCAGGAGAGACTGTCTGCTTGATT 361
      |||||||
QY 313 TyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLys 332
      |||||||
Db 362 TACGAGGACAGTGGATGACTTCCAGTGTGAATCAATCAATCAATCAATCAATCAATCA 421
      |||||||
QY 333 AspArgGluThrValLeuSerSerAlaLeu 342
      |||||||
Db 422 GAAGGAGGAGGAGTACCATCATCATATTA 451

RESULT 12
BQ891432      936 bp      mRNA      linear      EST 16-AUG-2002
LOCUS        AGENCOURT_8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
DEFINITION   5', mRNA sequence.
ACCESSION   BQ891432
VERSION     BQ891432.1 GI:22283446
KEYWORDS     EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgi.nci.nih.gov/.
            1 (bases 1 to 936)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            cDNA Library Preparation: ResGen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13790 row: n column: 19
            High quality sequence start: 25
            High quality sequence stop: 632.
FEATURES
            Location/Qualifiers
            1..936
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:6332754"
            /clone_lib="NIH_MGC_130"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb;
            Site:1: EcorV; Site:2: NotI; Cloned unidirectionally.
            Primer: Oligo dT. Average insert size 1.95 kb.
            Constructed by ResGen, Invitrogen Corp. Note: this is a
            NIH_MGC Library."
BASE COUNT   273 a 250 c 244 g 163 t 6 others
ORIGIN
Alignment Scores:
Pred. No.:      1.52e-36      Length:      936
Score:          696.50      Matches:    137
Percent Similarity: 91.7%      Conservative: 7

```


Best Local Similarity:	87.26%	Mismatches:	7
Query Match:	35.90%	Indels:	6
DB:	14	Gaps:	1

US-09-763-712A-2_COPY_206_547 (1-342) x BQ891432 (1-936)

QY	1	MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle	20
Db	476	ATGCAGCAAGACATGATGAGGTCAAAAGTTAGACACTGAAGTGCCTTATCAGTGTT	535
QY	21	MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhenThrIle	40
Db	536	ATGGAGAGATGAACATGGTTGACTCCAAGCAGCGTCAGCTCATCAAGAACTTTACCAT	595
QY	41	LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	60
Db	596	CTACAAGGTCTCTGGCCCCCAGAGTCCAAAAGGTGCAGAGGATCTCAGGAGCACCT	655
QY	61	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro	80
Db	656	GGTCCAACTGGCAACAGGAGACAAAGAGAGAGAGGAGAGCCTGGTCCACTGGCCCT	715
QY	81	AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly	100
Db	716	CGGGGTGAGAGGGCAATTTGACCAGTCTGGCCCTCCTGGAGAGCGTGGCAGCAAGGA	775
QY	101	SerLysGlySerGlnGly-ProLysGlySerArgGlySer-ProGlyLysProGlyProG	120
Db	776	TCCAAAAGGCTACAGGNTCCAAAGGGATCTGGTGGGTCCCGCAGGAAGCTGGCCCTC	835
QY	120	IleGlyProSerGlyAspProGly-ProProGlyProProGly-LysGluGlyLeuPro-G	139
Db	836	AAGGACTAGTGGGGACCCAGGNACCACCAAGTCCACCAAGNCAGGGATGGACTCCCTGG	895
QY	139	lyProGlnGlyProPro---GlyPheGlnGlyLeuGlnGly	151
Db	896	GCCTTCAGGNCNTCTGGGTCTCAGGGGACTACAAGGG	936

RESULT 13	
BMG76508/c	
LOCUS	500 bp mRNA linear EST 27-FEB-2002
DEFINITION	UI-E-EJO-ahq-c-05-0-UI.s2 UI-E-EJO Homo sapiens cDNA clone
	UI-E-EJO-ahq-c-05-0-UI 3', mRNA sequence.
ACCESSION	BMG76508
VERSION	BMG76508.1 GI:18986404
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 500)
AUTHORS	Ronald M.F., Lennon G. and Soares M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477

COMMENT	<p>Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu</p> <p>Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 Forward POLYA=Yes.</p>
FEATURES	Location/Qualifiers
source	1..500

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ahq-c-05-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site.1: EcoR I; Site.2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCCAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAGTC; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-EJ0
TAG_TISSUE=human eye anterior segment
TAG_SEQ=AATGCCGCAT"

```

BASE COUNT	109 a	123 c	89 g	179 t
ORIGIN				

Alignment Scores:						
Pred. No.:	5.6e-35	Length:	500			
Score:	669.00	Matches:	117			
Percent Similarity:	100.0%	Conservative:	0			
Best Local Similarity:	99.1%	Mismatches:	0			
Query Match:	34.48%	Indels:	0			
DR:	14	Gaps:	0			

US-09-763-712A-2_COPY_206_547 (1-342) x BM676508 (1-500)

Qy	225	Lys	Glu	Ile	Phe	Glu	Asp	Ala	Lys	Leu	Phe	Cys	Glu	Asp	Lys	Ser	His	Leu	Val	Phe	244
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	3
Db	495	GAG	AAA	TTTTT	TGG	ATG	CA	AG	CTTT	TCT	G	A	CA	CA	AG	CTCT	C	A	C	ATCT	436
Qy	245	Ile	Asn	Thr	Arg	Glu	Glu	Gln	Gln	Trp	Ile	Leu	Val	Lys	Gln	Met	Val	Glu	Arg	Glu	264

435 ATAAACACTAGAGAGGACAGCATGGATAAAAACACATGGTAGGACAGAGAGCCAC 376

[illegible][illegible]

DD 3/3 TGGATCGGGCCTCACAGACTCAGAGCGGTGAAATGAAATGGAGTGGCTGGATGGGACATCT 31

QY 285 proAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyPro 304

Db 315 CCAGACTACAAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCCATGGGCCA 256

Qy 305 GlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTyrAsnAspPheGlnCysGluAsp 324

Db 255 GGAGAAGACTGTGCTGGGTTGATTATGCTGGGCAGTGGACGATTTCCAATGTGAAGAC 196

Qy 325 ValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerAlaLeu 342

Db 195 GTCAATAACTTCAATTGCGAAAAGACAGGAGACAGTACTGTCATCTGCATTA 142

RESULT 14

BM713891
LOCUS

DEFINITION

ACCESSION
VERSION

BM713891 507 bp mRNA linear EST 28-FEB-2002
 UI-E-EJ0-ahq-c-05-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ahq-c-05-0-UI 5', mRNA sequence.

BM713891
BM713891 1 GT:19027149

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84


```

Db      25  CCAGTGGTCCTTGGGGCCCTGGAATAATGGAATCTTGGGCTTCTGGTCCGCCAGGA 84
QY      146  PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu--- 164
Db      85  CCACAGGGGAAAGAGGATCAATGGGAACTGGAGCTCCAGGGATTAGAGGGTCATCT 144
QY      165  -----ProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyPro 182
Db      145  GGACCTCCAGGAATACCTGGATTACCTGGAAATACCTGGACCTAGAGGTCTCCAGGGCCT 204
QY      183  ProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro----- 200
Db      205  CCTGGAGATTGGGAAGT-----TCTTTAGCATTACATCAGTCTCTTCACAGAGAAGTA 258
QY      201  -----AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCys 218
Db      259  ATGCTTGAACCTGTGGTGACCGGTGCCCTCTGGAGTGGGAAGAACTTTCTGATAAATGC 318
QY      219  TyrTyrPheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLys 238
Db      319  TACTACTTCTCAACTGGAAGAGATATATTGATGATGCCAAACTGATCTGTAAGAAAAG 378
QY      239  SerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMet 258
Db      379  GCAGCGATGTTAGTTGTGATTGAACCCCAAGCAGAGCAACAATTTTGAAGAAACAACA 438
QY      259  ValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 278
Db      439  AGTGGAAAGGAACACTTCTGGATAGTCTCAGAGATCTGAGGAGAGGGGTACATGGAAA 498
QY      279  TrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrp 298
Db      499  TGGCTGGATGGTACTACTCTAAACTACAAAAAAGTGAAGAGAGGACAGCCGGACGATTGG 558
QY      299  GlyHisGlyHisGlyProGlyGluAspCysAla 309
Db      559  TCTCATCACACTGGACCAGGAGAGGATTGCGCT 591

```

Search completed: March 21, 2003, 12:20:33
 Job time : 1136.43 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run On: March 21, 2003, 02:20:06 ; Search time 239.704 Seconds
(without alignments)
4293.470 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547
Perfect score: 2520
Sequence: 1 MNSOLNFTGOMENTTISQ.....EDVNFNICEKDRVLSAL 457

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spoal/US09763712/runat_14032003_100949_18091/app_query.fasta_1.1877
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.ecgn.1.1.626.e@unat.14032003_100949_18091 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2520	100.0	2024	21	AAA07697 Human collectin en
2	2520	100.0	2262	24	ABA97932 Human scavenger re
3	2520	100.0	2628	22	AAH43036 Nucleotide sequenc
4	2514	99.8	2181	22	AAI60628 Human polynucleoti
5	2514	99.8	2641	22	AAC66903 Human EXNAD-14 cod
6	2514	99.8	2929	24	ABQ92072 Human polynucleoti
7	2514	99.8	2930	20	AAV55746 Human secreted pro
8	2507	99.5	2318	22	AAI58842 Human angiotensin
9	2487	98.7	2005	24	ABL95574 Human polynucleoti
10	2487	98.7	2005	24	ABL88085 Human PRO7223 CDNA
11	2344	93.0	2637	22	AAH43037 Nucleotide sequenc
12	1738	69.0	2256	22	AAH43054 Nucleotide sequenc
13	1715.5	68.1	1521	23	AAH71133 DNA encoding novel
14	682	27.1	3685	20	AAH27858 Human CSR3 protein
15	682	27.1	3810	20	AAH27856 Human CSR1 protein
16	474.5	18.8	1410	24	ABK84622 Human CDNA differe
17	474.5	18.8	1410	24	ABL66964 Lung cancer relate
18	471	18.7	5676	15	AAQ64556 Human collagen (Ty
19	471	18.7	8284	22	AAH26526 Human breast cance
20	471	18.7	8284	22	AAH26553 Human breast cance
21	471	18.7	8284	22	AAH26600 Human breast cance
22	462	18.3	4428	22	AAH06574 Bovine alpha1(III)
23	462	18.3	4428	22	AAH06575 Bovine alpha1(III)
24	459.5	18.2	756	14	AAQ43034 Collagen-like poly
25	459.5	18.2	756	17	AAH16768 Collagen-like poly
26	458.5	18.2	4821	23	AAH86866 DNA encoding novel
27	454	18.0	3171	21	AAH12503 DNA encoding a hu
28	454	18.0	3174	21	AAH29843 DNA encoding huma
29	453	18.0	6158	24	ABL62095 Colon adenocarcino
30	453	18.0	6158	24	ABL65452 Lung cancer relate
31	453	18.0	6158	24	ABK3486 Human endometrial
32	452.5	18.0	3170	21	AAH12493 CDNA encoding huma
33	452.5	18.0	3171	21	AAH12502 CDNA encoding a hu
34	452.5	18.0	3181	19	AAV59358 Nucleotide sequenc
35	452.5	18.0	3349	17	AAT16518 Collagen A1/decori
36	452.5	18.0	3349	21	AAH12500 CDNA encoding a ch
37	452.5	18.0	3531	21	AAH12497 CDNA encoding a ch
38	452.5	18.0	3535	17	AAH16515 Collagen A1/BMP-2B
39	452.5	18.0	3541	17	AAH16516 Collagen A1/TGF-be
40	452.5	18.0	3541	21	AAH12498 CDNA encoding a ch
41	452.5	18.0	4192	17	AAT16517 Collagen A1/decori
42	452.5	18.0	4409	19	AAV60814 Human recombinant
43	452.5	18.0	4770	22	AAH22441 Human CDNA encodin
44	452.5	18.0	6728	22	AAF90491 Human pro-alpha-1
45	452.5	18.0	6728	24	ABN97451 Gene #3949 used to

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX
AC AAA07697;
XX
XX 20-JUN-2000 (first entry)
DT
DE Human collectin encoding DNA.
DE
XX Collectin; human; antibacterial; antiviral; ds.
KW
OS Homo sapiens.
XX
XX Location/Qualifiers
FT CDS
FT 55..1698
FT /*tag= a
FT /product= "collectin"


```
QY 381 lleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
Db 1465 ATCGCCCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGCTGGATGGGACATCTCCA 1524
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
Db 1525 GACTACAAAATTTGGAAGCTGGACAGCCGATAACTGGGTTCATGGCCATGGGCCAGGA 1584
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 1585 GAAGACTGCTGGGTTGATTTATGCTGGCAGTGGAGCAGGATTTCCATGTGAAGACGTC 1644
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
Db 1645 AATRACTTTCAATTTGGGAAAAAGACAGGGGAGAGCAGTACTGTCTATCTGCATTA 1695
RESULT 2
ABA97932
ID ABA97932 standard; cDNA; 2262 BP.
XX
AC ABA97932;
XX
DT 25-APR-2002 (first entry)
XX
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
KW Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
KW rheumatoid arthritis; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 33..2262
FT /tag= a
FT /product= "scavenger receptor-like protein"
XX
PN JP2001340089-A.
XX
PD 11-DEC-2001.
XX
PF 08-DEC-2000; 2000JP-0375066.
XX
PR 27-MAR-2000; 2000JP-0090772.
XX
PA (SHIO ) SHIONOGI & CO LTD.
XX
WPI: 2002-144965/19.
DR P-PSDB; ABB08642.
XX
PT New scavenger receptor-like protein for diagnosis, prevention and
PT treatment of autoimmune disease, such as rheumatoid arthritis -
XX
PS Claim 3; Fig 1; 38pp; Japanese.
XX
CC The invention relates to a human scavenger receptor-like protein. The
CC protein is useful as a target molecule for diagnosis, prevention and
CC treatment of autoimmune diseases such as rheumatoid arthritis.
XX
SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
Alignment Scores:
Pred. No.: 1.9e-80 Length: 2262
Score: 2520.00 Matches: 457
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0
US-09-763-712A-2_COPY_91_547 (1-457) x ABA97932 (1-2262)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIleSerGln 20
Db 889 ATGAACAGCCAGCTCAACTATTTCACAGTTCAGATGGAGAACATCACCACCTATCTCTCA 948
```

```
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 949 GCCAACGAGCAGAACTCTCAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAAGACA 1008
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 1009 GCCATCAAGTTTCAACCAACTGGAGGAGCGTTTCCAGCTCTTTTGAGACGGATATTGTGAAC 1068
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuAspThrLeuThrSerAsnLeuAsn 80
Db 1069 ATCATTAGCAATATCAGTTTACACAGCCACACCTGCGGACGCTGACGCAATCTAAT 1128
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 1129 GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACACAGATGATCTGACCTCCTTG 1188
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 1189 AATAATACCTGGCCAAACATCCGTTTGGATTCTGTCTCTCAGGATGCAACAAGATTG 1248
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
Db 1249 ATGAGGTCGAGGTTAGACACTGAAGTAGCAACTTATCAGTGATTATGGAAGAAATGAAG 1308
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
Db 1309 CTAGTAGACTCCAAGCATGGTCAGCTCATCAAGATTTTACANTACTACAGGTCCACCG 1368
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
Db 1369 GGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCTGGCCCAACTGGCAAC 1428
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
Db 1429 AAGGACACAAAGGAGAGAGAGGGGAGGCTGGACCACCTGGCCCTGCGGGTGAGAGAGGC 1488
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
Db 1489 CCAATTGGACCACTGGTCCCTCCCGAGAGCGTGGCGCAAGAGATCTAAAGGCTCCAG 1548
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
Db 1549 GGCCCAAAAGGCTCCCGTGGTTCCCTGGAAAGCCGCGCTCAGGGGCCCAAGTGGGAC 1608
QY 241 ProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProProGly 260
Db 1609 CCAGGCCCCCGGGCCCAAGAGGAGGACTCCCGGGCCCTCAGGGCCCTCCTGGC 1668
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1669 TTCCAGGGACTTCAGGGCACCGTTGGGAGCGCTGGGTGCTGGACCTCGGGGACTGCCA 1728
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProProGlyPro 300
Db 1729 GGCTTGCTGGGGTACCAGGATGCCAGGCCCAAGGGCCCCCGGGCCCTCCTGGCCCA 1788
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1789 TCAGGAGCGGTGGTCCCTCGCCCTGCAGATGAGCAACCCCGGACCGGAGGACAAAT 1848
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
Db 1849 GGCTGCCCGCTCACTCGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAA 1908
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
Db 1909 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTTTTCATA 1968
QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
Db 1969 AACACTAGAGGAGAACGCAATGGATAAAAAAGATGGTGGTGGAGAGAGAGGACCTGG 2028
```

Qy	381	IleClyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro	400
Db	2029	ATGGGCTCACAGACTCAGAGCGTGAATAATGAATGAAAGTGGCTGGATGGGCACATCTCCA	2088
Qy	401	AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly	420
Db	2089	GACTACAAAAATTGGAAAGCTGGACACGCGGATAACTGGGTCATGGCCATGGGCCAGGA	2148
Qy	421	GluAspCysAlaGlyLeuIleValAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal	440
Db	2149	GAAGACTGTGCTGGGTGATTATGCTGGCAGTGGACGATTCCATGTGAAGAGTC	2208
Qy	441	AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu	457
Db	2209	AATAACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATTA	2259
RESULT 3			
AAH43036			
ID	AAH43036 standard; DNA; 2628 BP.		
XX			
AC	AAH43036;		
XX			
DT	15-OCT-2001 (first entry)		
XX			
DE	Nucleotide sequence of a human scavenger receptor.		
XX			
KW	Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;		
KW	arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;		
KW	low density lipoprotein; ss.		
XX			
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
FT	74..2302		
CDS	/*tag= a		
FT	/product= "scavenger receptor"		
FT			
XX			
XX	WC200159107-A1.		
PN			
XX			
PD	16-AUG-2001.		
XX			
PF	08-FEB-2001; 2001WO-JP00874.		
XX			
XX	14-FEB-2000; 2000JP-0035155.		
PR	10-OCT-2000; 2000JP-0309068.		
XX			
PA	(FUSO) FUSO PHARM IND LTD.		
XX			
PI	Wakamiya N;		
XX			
XX	WPI; 2001-497076/54.		
DR	P-PSDB; AAG63346.		
DR			
XX			
PT	New Scavenger receptor proteins SRCL-P1 with collectin-like structure,		
PT	useful for treatment and diagnosis of diseases associated with oxidized		
PT	low-density lipoprotein accumulation -		
XX			
PS	Claim 2; Page 79-84; 118pp; Japanese.		
XX			
CC	The present sequence encodes a human scavenger receptor, designated		
CC	SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They		
CC	are useful in clarifying the functions of macrophages and basal		
CC	immunity. They are also useful in the treatment, prevention, diagnosis		
CC	and investigation of diseases such as arteriosclerosis, diabetic		
CC	complications, bacterial infection and restenosis following angioplasty,		
CC	which are associated with accumulation of oxidized low density		
CC	lipoprotein and the binding of advanced glycation end-products into		
CC	cells.		
XX			
SQ	Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;		
Alignment Scores:			
Pred. No.:	2.13e-80	Length:	2628

Score:	2520.00	Matches:	457
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0
US-09-763-712A-2_COPY_91_547 (1-457) x AAH43036 (1-2628)			
Qy	1	MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln	20
Db	929	ATGAACAGCCAGCTCAACTATTTCACAGGTCAAGTGGAGACATCCACCACATCTCTCTCAA	988
Qy	21	AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr	40
Db	989	GCCAAACGAGCAGAACCTGAAAGACCTGCAGGACTTACACAAGATGCAGAGAATAGAACA	1048
Qy	41	AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn	60
Db	1049	GCCATCAAGTTCACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC	1108
Qy	61	IleIleSerAsnIleSerTyThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn	80
Db	1109	ATCATTAGCAATATCATGTACAGCCACCACTTGGGAGCGTGCAGCAATCTAAAT	1168
Qy	81	GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu	100
Db	1169	GAAGTCAGGACCATTGTCACAGATACCCTTACCAACACACAGATGATCTCACCTCTTG	1228
Qy	101	AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu	120
Db	1229	AATAATACCTGGCCAACTCCGTTTGGATTCTGTCTTCAGGATGCAACAAGATTG	1288
Qy	121	MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys	140
Db	1289	ATGAGTCTGAGGTTAGACACTGAAGTAGCAACTATTACAGTGATTATGGAAGAAATGAAG	1348
Qy	141	LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro	160
Db	1349	CTAGTAGACTCCAAGCATGTCAGCTCATCAAGAATTTTACAATACTACAAGGTCCACCG	1408
Qy	161	GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsn	180
Db	1409	GGCCCCAAGGGTCCAAGAGTGACAGAGATCCCAAGGACCCCTGGGCCAACTGGCAAC	1468
Qy	181	LysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGluArgGly	200
Db	1469	AAGGCACGAAGAGAGAGAAGGGGAGCCTGGACCACTGGCCCTGCGGGTGCAGAGAGGC	1528
Qy	201	ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln	220
Db	1529	CCAATTGCACAGCTGCTCCCCCGGAGAGCGTGGCGCAAAAGGATCTAAAGGCTCCACG	1588
Qy	221	GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp	240
Db	1589	GGCCCCAAGGCTCCCGTGGTTCCTCGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGAC	1648
Qy	241	ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly	260
Db	1649	CCAGGCCCCCGGGCCACACAGGCAAGAGGACATCCCCGGCCCTCAGGGCCCTTCTGGC	1708
Qy	261	PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro	280
Db	1709	TTCCAGGCACTTCAGGGCACCGTTGGGAGCCCTGGGTCCTGGACCTCGGGAGTCGCA	1768
Qy	281	GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro	300
Db	1769	GGCTTGGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGGGCCCA	1828
Qy	301	SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn	320
Db	1829	TCAGAGCGGTGTGGCCCTTGCAGATAGCCAAACCCCGGACCCGGAGGACAT	1888
Qy	321	GlyCysProProHisTrpLysAsnPheThrAspLysCysTyTrpPheSerValGluLys	340


```
Db 917 CCAATTGGACAGCTGGTCCCTCCCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCAG 976
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
Db 977 GGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCAGGGCCCGAGGGGAC 1036
QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
Db 1037 CCAGGCCCCCGGCGCCACAGGCAAGAGGACTCCCGGCCCTCAGGGCCCTCCTGGC 1096
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyValProArgGlyLeuPro 280
Db 1097 TTCAGGAGCTTCAGGACCCCTGGGAGCCCTGGGGTGCCTGGACTCGGGAGCTGCCA 1156
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
Db 1157 GGCTTGGCTGGGTACCAAGCATGCCAGGCCCAAGGCCCGCCCGCCCTCCTGGCCCA 1216
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1217 TCAGGAGCGGTGGTCCCTGGCCCTGCAGATGAGCCCAACCCCGCACCGGAGACAAT 1276
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyrPheSerValGluLys 340
Db 1277 AGCTGCCGCTCCTGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 1336
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
Db 1337 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGCTTCACATCTTGTTCATA 1396
QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
Db 1397 AACACTAGAGAGAGCAAGCAATGATAAAAAACAGATGCTAGGGAGAGAGCCACTGG 1456
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
Db 1457 ATCGGCTCACAGACTCAGACGCTGAAATGAATGAAGTGGCTGGTGGGATCTCCA 1516
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
Db 1517 GACTACAAAAATTGGAAGCTGGACAGCCGGATACTGGGTGATGCCATGGCCAGGA 1576
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 1577 GAAGACTGTGCTGGGTGATTTATGCTGGCGAGTGGAGCAATTTCCAATGTGAAGACGTC 1636
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
Db 1637 AATACTTCATTTGCCGAAAAAGACAGGGAGACAGTACTGTCTCATCTCATTA 1687
RESULT 5
AAC66903
ID AAC66903 standard; cDNA; 2641 BP.
XX
AC AAC66903;
XX
27-MAR-2001 (first entry)
DT
DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
DE
KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
KW inflammation; reproductive disorder; cardiovascular disorder;
KW immune disorder; musculoskeletal disorder; developmental disorder;
KW gastrointestinal disorder; cell proliferation disorder; ss.
OS Homo sapiens.
XX
PN WO200068380-A2.
XX
PD 16-NOV-2000.
XX
PF 10-MAY-2000; 2000WO-US12811.
XX
```

```
PR 11-MAY-1999; 99US-0133643.
PR 23-AUG-1999; 99US-0150409.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
PI Azimzai Y;
XX
DR WPI; 2001-007395/01.
DR P-PSDB; AAB27236.
XX
Isolated polynucleotide encoding extracellular matrix or
PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT preventing disorders associated with expression of EXMAD such as
PT proliferative, immune and genetic disorders -
XX
PS Claim 4; Page 121-122; 129pp; English.
XX
The present invention provides the protein and coding sequences for 25
CC novel extracellular matrix and adhesion-associated proteins (EXMADS).
CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC useful in the prevention and treatment of cancers, cell proliferation,
CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC gastrointestinal disorders and inflammation.
XX
SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;
Alignment Scores:
Pred. No.: 3.46e-80 Length: 2641
Score: 2514.00 Matches: 456
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: 22 Gaps: 0
US-09-763-712A-2_COPY_91_547 (1-457) x AAC66903 (1-2641)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 567 ATGAACAGCGCAGCTCAACTCACTTCACAGGTTCAGATGGAGAACATCACCATCTCTCTCAA 626
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 627 GCCAACGACGAGAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGATAGACA 686
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 687 GCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATGTGAAC 746
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
Db 747 ATCATTTAGCAATATCATGTTACACAGCCCACTCCGCGCTGCGGACCAATCTAAAT 806
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 807 GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 866
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 867 AATAATACCTCGCCCAACATCCGTTTGGATTCTCTCTCAGGATGCAACAAGATTG 926
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
Db 927 ATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTATATGAAGAAATGAAG 986
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
Db 987 CTAGTAGACTCCCAAGCATGGTCAGCTCATCAAGATTTTACATACTACAGGTCCACCG 1046
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
```


[illegible]

KW	immunostimulant; antiparasitic; secreted protein; transmembrane protein; cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW	stem cell; growth factor; nervous system disease; neuropathy;
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW	osteoporosis; severe combined immunodeficiency; SCID; infection;
KW	multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	US2002065394-A1.
PN	
PD	30-MAY-2002.
XX	
PF	22-DEC-2000; 2000US-0745763.
XX	
PR	18-MAR-1998; 98US-0040963.
XX	
PA	(JACO/) JACOBS K.
PA	(MCCO/) MCCOY J M.
PA	(LAVA/) LAVALLIE E R.
PA	(COLL/) COLLINS-RACIE L A.
PA	(EVAN/) EVANS C.
PA	(MERB/) MERBERG D.
PA	(TREA/) TREACY M.
PA	(SPAU/) SPAULDING V.
XX	
PI	Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI	Merberg D, Treacy M, Spaulding V;
XX	
DR	WPI: 2002-582343/62.
DR	P-PSDB: ABP61859.
XX	
PT	Novel secreted or transmembrane protein and polynucleotide encoding the
PT	protein, useful for diagnosis and treatment of neurological disorders,
PT	cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
XX	
PS	Claim 250; Page 231-232; 284pp; English.
XX	
CC	The invention relates to human secreted or transmembrane protein (I),
CC	their fragments and is encoded by specific complementary deoxyribonucleic
CC	acid (cDNA) inserts (II), where the protein is substantially free from
CC	other mammalian proteins. (I) are useful for preventing, treating or
CC	ameliorating a medical condition, especially immunological treatment or
CC	prevention of tumours. (I) exhibits activity relating to angiogenesis,
CC	cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC	stem cell growth factor activity and activin or inhibin-related
CC	activities. (I) can be used to manipulate stem cells in culture to give
CC	rise to neuroepithelial cells that can be used to augment or replace
CC	cells damaged by illness, autoimmune disease, accidental damage or
CC	genetic disorders. (I) induces the proliferation of neural cells and
CC	regeneration of nerve and brain tissue and is useful for the treatment of
CC	central and peripheral nervous system diseases and neuropathies, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC	activity, regulation of haematopoiesis and is useful for treating myeloid
CC	or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC	and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC	tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC	for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC	periodontal disease. (I) is also useful for gut protection or
CC	regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC	in various tissues, various immune deficiencies and disorders including,
CC	severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC	autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC	diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC	such as asthma or other respiratory problems. (II) is useful to express
CC	recombinant protein, as markers for tissues in which the corresponding
CC	protein is preferentially expressed and in gene therapy. The present
CC	sequence is that of a polynucleotide of the invention.
XX	
SQ	Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:

Pred. No.: 3,74e-80 Length: 2929
 Score: 2514.00 Matches: 456
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 24 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x ABQ92072 (1-2929)

QY	1	MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIleSerGln	20
DB	921	ATGAACAGCCAGCTCAACTCATTCACAGGTTCAGATGGAGAACATCACCATATCTCTCAA	980
QY	21	AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr	40
DB	981	GCCACGAGCAGAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGATAGAACA	1040
QY	41	AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn	60
DB	1041	GCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC	1100
QY	61	IleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn	80
DB	1101	ATCATTAGCAATATCAGTTACACAGCCACCACTCGGACGCTGACCAAGCAATCTAAAT	1160
QY	81	GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu	100
DB	1161	GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTGG	1220
QY	101	AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu	120
DB	1221	AATAATACCTGGCCCAACATCCCTTGGATTCTGTCTCTCAGGATGCAACAAGATTG	1280
QY	121	MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys	140
DB	1281	ATGAGTCGAGGTATGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG	1340
QY	141	LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro	160
DB	1341	CTAGTAGACTCCCAAGCATGTGCTAGCTCATCAAGAAATTTACATACTACAAGTCCACCG	1400
QY	161	GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn	180
DB	1401	GGCCCCAGGGGTCACAGAGGTGACAGAGGATCCAGGACCCCTGCGCCCACTGGCAAC	1460
QY	181	LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly	200
DB	1461	AAGGCACAGAAAGAGAGAGAGGGGGGACCTGGACCACTGGCCCTCGGGTGAGAGGC	1520
QY	201	ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerGlySerGln	220
DB	1521	CCAATTGGACCACTGGTCCCTCCCGGAGAGCGGTGGCGCAAGGATCTAAAGGCTCCAG	1580
QY	221	GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp	240
DB	1581	GGCCCCAAAGGCTCCGCTGGTTCCCTTGGGAAGCCCGGCCCTCAGGGCCCTCTGGC	1640
QY	241	ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly	260
DB	1641	CCAGGCCCCCGGGCCCAACAGCAAGAGGGACTCCCGGCCCTCAGGGCCCTCTGGC	1700
QY	261	PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro	280
DB	1701	TTCAGGGACTTCAGGACCCGTTGGGAGCCCTGGGTGCTGGACCTCGGGGACTGCCA	1760
QY	281	GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro	300
DB	1761	GGCTTCCTGGGTATACAGGCATGCGAGGCCCCCAAGGCCCTCCCGGCCCTCTGGCCCA	1820
QY	301	SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn	320
DB	1821	TCAGGAGCGGTGTGGCCCTTGGCCCTTGCAGATGACCAACCCCGGACCGGAGACAAT	1880

QY	321	GlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyrPheSerValGluLys	340
DB	1881	AGCTGCCCGCTCCTACCTGGAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA	1940
QY	341	GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle	360
DB	1941	GAAATTTTTCAGGATGCAAGCTTTCTGTGGAAGCAAGTCTTCACATCTGTGTTTCATA	2000
QY	361	AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp	380
DB	2001	AACACTAGAGAGAAACAGCAATGGATAAAAAACAGATGTAGGAGAGAGACCCACTGG	2060
QY	381	IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro	400
DB	2061	ATCGGCTCACAGACTCAGAGCGTGAAATGAATGAAGTGGCTGGATGGGACATCTCCA	2120
QY	401	AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyGlyProGly	420
DB	2121	GACTACAAAAATTTGGAAGCTGGACACCGCGATAACTGGGGTCATGCCCATGGCCAGGA	2180
QY	421	GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal	440
DB	2181	GAAGACTGTCTGGGTGATTATGCTGGCAGCTGGAAACGATTCCAATGTGAAGCGTC	2240
QY	441	AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu	457
DB	2241	AATAACTTCATTGGCAAAAAAGACAGGACAGACAGTACTGTCTCATCTGCATTA	2291

RESULT 7

AAV55746

ID AAV55746 standard; cdNA; 2930 BP.

AC AAV55746;

DT 23-MAR-1999 (first entry)

DE Human secreted protein clone bv227_1 coding sequence.

Secreted protein; human; nutritional supplements; immune stimulant;
 immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
 activin/inhibin; chemokinesis; haemostasis; thrombolysis;
 receptor/ligand activity; anti-inflammatory; tumour inhibitor;
 cadherin/tumour invasion suppressor; ds.

OS Homo sapiens.

Key Location/Qualifiers

CDS 67..693

/*tag= a

XX WO9855614-A2.

XX 10-DEC-1998.

XX 01-JUN-1998; 98WO-US11210.

XX 29-MAY-1998; 98US-0087255.

XX 04-JUN-1997; 97US-0868696.

XX 04-JUN-1997; 97US-0868697.

XX 04-JUN-1997; 97US-0868698.

XX 04-JUN-1997; 97US-0868699.

XX 04-JUN-1997; 97US-0868900.

XX 04-JUN-1997; 97US-0869191.

XX 04-JUN-1997; 97US-0869192.

XX 04-JUN-1997; 97US-0869193.

XX 04-JUN-1997; 97US-0869194.

PA (GEMY) GENETICS INST INC.

PI Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
 McCoy JM, Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-059912/05.
 DR P-PSDB; AAW73628.
 XX
 PT New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 XX
 PS Claim 26: Page 87-88; 127pp; English.
 XX
 CC This sequence encodes a human secreted protein of the invention.
 CC This DNA sequence was isolated from a human adult brain cDNA
 CC library, and was designated clone bv227-1. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in
 CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 XX
 SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:
 Pred. No.: 3,74e-80 Length: 2930
 Score: 2514.00 Matches: 456
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 20 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x AAV55746 (1-2930)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIleSerGln 20
 DB 921 ATGAACAGCCAGCTCAACTCATTACAGGTGAGTGGAGAACATCACCCTATCTCTCAA 980
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 DB 981 GCCAAGCAGCAGCACTGGAAGACCTGCGAGGACTTACACAAAGATGCAGAGAATAAGAACA 1040
 QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 DB 1041 GCCATCAAGTTCACCAACCTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 1100
 QY 61 IleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 DB 1101 ATCATTAGCAATATACATTACAGCCCAACCACTGCGGACGCTGACCAAGCAATCTAAAT 1160
 QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 DB 1161 GAAGTCAGGACCACTTGCCAGATACCCCTTACCAACACACAGATGATCTGACCTCCTTG 1220
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnAspLeu 120
 DB 1221 AATAATACCTGGCCCAACATCGGTTGGATTCTGTCTCTCAGGATGCAACAAGATTGG 1280
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 DB 1281 ATGAGGTCAGGTTAGACACTGAGTAGCCAACTTATCATGATATTGGAAGAAATGAAG 1340
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 DB 1341 CTAGTAGACTCCAAAGCATGGTCAGTCAATCAAGAATTTTACAATACTACAAGGTCCACCG 1400
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
 DB 1401 GGCCCCCAGGGGTCCAAGAGGTGACAGAGATCCAGGGGACCCCTGGCCCCAACATGGCAAC 1460

QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
 DB 1461 AAGGGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1520
 QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 DB 1521 CCAATTGGACACAGCTGGTCCCGGAGAGCGTGGCGCAAGAGATCTAAAGGCTCCAG 1580
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 DB 1581 GCGCCCAAGAGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGAC 1640
 QY 241 ProGlyProProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProProGly 260
 DB 1641 CCAGGCCCCCGGCCCCACAGCAAGAGGAGCTCCCGGCCCTCAGGGCCCCCTCTGGC 1700
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 DB 1701 TTCCAGGGGACTTCAGGGGACCGTTGGGAGCGCTGGGCTGCTGGACCTCGGGACTGCCA 1760
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
 DB 1761 GGTTCCTGGGTACAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCTCTCTGGCCCCA 1820
 QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 DB 1821 TCAGGAGCGGTGGTCCCTCGCCCTGCAGAAATGAGCAACCCCGCCAGGAGACAAT 1880
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTrpPheSerValGluLys 340
 DB 1881 AGCTGCCCGCTCACTTGAAGAACTTCACAGACAAATGTACTATTTTTCAGTTGAGAAA 1940
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 DB 1941 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATA 2000
 QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 DB 2001 AACACTAGAGAGAACAGCAATGGATAAANAACACATGGTAGGAGAGAGAGAGAGAG 2060
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 DB 2061 ATCGGCTCACAGACTCAGAGCGTGAAATGAATGAAGTGGCTGGATGGGACATCTCCA 2120
 QY 401 AspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyGlyProGly 420
 DB 2121 GACTCAAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGGCCAGGA 2180
 QY 421 GluAspCysAlaGlyLeuIleTrpAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 DB 2181 GAAGACTGTGCTGGGTGATTTATGCTGGCAGTGAACGATTTCCATGTAAGACGTC 2240
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 DB 2241 AATAACTTCAATTTGCAAAAAGACAGGAGAGACAGTACTGTCTCATCTGCATTA 2291

RESULT 8

AAI58842

ID AAI58842 standard; cDNA; 2318 BP.

XX

AC AAI58842;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 1045.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB; AAM39886.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 1045; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,52e-80 Length: 2318
 Score: 2507.00 Matches: 457
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 0
 Query Match: 99.48% Indels: 1
 DB: 22 Gaps: 0
 US-09-763-712A-2_COPY_91_547 (1-457) x AAI58842 (1-2318)
 QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB 317 ATGAACAGCGAGCTCAACTCAATTCACAGGTACAGGTGAGAGAACATCACCCTCTCTCAA 376
 QY 21 -AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgTh 40
 DB 377 GCCCAACGAGCAGACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAGAAC 436
 QY 40 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAs 60
 DB 437 ACCCATCAAGTTCAACCACTGGAGGACGCTTCCAGCTCTTTGAGCGGATATTGTGAA 496

QY 60 nIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAs 80
 DB 497 CATCATTTAGCAATATCAGTTACACAGCCACACCTGCGAGCGCTGACCAGCAATCTAAA 556
 QY 80 nGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLe 100
 DB 557 TGAAGTCAGGACCACTTGCACAGATACCTTACCAAAACACACAGATGATGTGACCTT 616
 QY 100 uAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLe 120
 DB 617 GAATATATACCTTGGCCCAACATCCGTTTGGATTTCTGTTCTCTCAGGATGCAACAAGATTT 676
 QY 120 uMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetIly 140
 DB 677 GATAGGTCGAGTTAGACACTGAAGTAGCAACTTATCAGTATGATGAGAGAAATGAA 736
 QY 140 sLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPr 160
 DB 737 GCTAGTAGACTCCAAGCATGGTCACTCATCAAGAAATTTTACAATATCTACAAGGTCCACC 796
 QY 160 oGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAs 180
 DB 797 GGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCCAAGGACCCCTGCCCCCACTGGCAA 856
 QY 180 nLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGluArgG 200
 DB 857 CAAGGACACAAAGGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTGCGGGTGAGAGAG 916
 QY 200 yProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerG 220
 DB 917 CCCAATTTGACCACTGGTCCCGCCGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCCA 976
 QY 220 nGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAs 240
 DB 977 GGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGAGAGCCCGCCCTCAGGGCCCCAGTGGGA 1036
 QY 240 pProGlyProProGlyProProGlyLysGlyGluGlyLeuProGlyProGlnGlyProProG 260
 DB 1037 CCCAGGCCCCCGGGGCCCAACAGGAGGAGGACATCCCGCGGCCCTCAGGGCCCTCTGTG 1096
 QY 260 yPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPr 280
 DB 1097 CTTCCAGGAGACTTCAGGGCACCCTTGGGAGCGCTGGGTGCTTGGACCTCGGGGACTGCC 1156
 QY 280 oGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPr 300
 DB 1157 AGGCTTGCCTGGGTACCAAGGATGCCAGGCCCAAGGCCCCCGCGGCCCTCTCTGGCCC 1216
 QY 300 oSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAs 320
 DB 1217 ATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCCAACCCCGGACACCGAGGACAA 1276
 QY 320 nGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLy 340
 DB 1277 TGGCTGCGCGCCTCACTTGAAGAATCTACAGACAAATGCTACTATTTTTCAGTTAGAGAA 1336
 QY 340 sGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheI 360
 DB 1337 AGAAATTTTGAAGATGCAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTGTTTTCAT 1396
 QY 360 eAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTr 380
 DB 1397 AAACACTAGAGAGGAACAGCAATGGATAAAAAAACAGATGGTAGGAGAGAGAGAGCCAC 1456
 QY 380 pIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPr 400
 DB 1457 GATCGGCCTCACAGACTCAGAGCGTGAAACCGAAATGGAAAGTGGCTGGATGGGACATCTCC 1516
 QY 400 oAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGl 420
 DB 1517 AGACTACAAAATTTGAAAGCTGGACCGCGGATAACTGGGGTCTATGGCCATGGGCCAGG 1576
 QY 420 yGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVa 440

```

|||||
Db 1577 AGAAGACTGTGGTTCATTATGCTGGCAGTGGACGATTTCCATGTGAAGACGT 1636
Qy 440 lAsnAsnPhelIeCysGLuLysAspArgGLuThrValLeuSerSerAlaLeu 457
|||||
Db 1637 CAATACCTTCATTTCGGRAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1688

RESULT 9
ABL95574
ID ABL95574 standard; cDNA: 2005 BP.
XX
AC ABL95574;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.
XX
KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW carotid; cytostatic; antiangiogenic; hypotensive; vulnerary;
KW antiarteriosclerotic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200208284-A2.
XX
PD 31-JAN-2002.
XX
PF 09-JUL-2001; 2001WO-US21735.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX
PA (GETH ) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERR/) GERRITSEN M E.

```

```

PA (GODD/) GODDARD A.
PA (GODO/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAON/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WL, Ye W;
XX
DR WPI: 2002-171999/22.
DR P-PSDB: ABB95436.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
XX Claim 1; Fig 27; 567pp; English.
XX
PS The present invention provides the protein and coding sequences of human
PS PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac
CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial stenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
CC healing. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Alignment Scores:
Pred. No.: 2.48e-79 Length: 2005
Score: 2487.00 Matches: 455
Percent Similarity: 99.56% Conservative: 0
Best Local Similarity: 99.56% Mismatches: 2
Query Match: 98.69% Indels: 1
DB: 24 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x ABL95574 (1-2005)

Qy 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 134 ATGAACAGCCAGCTCAACTCATTTCAGGTGAGATGGAGAACATCACCCTATCTCTCAA 193
Qy 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 194 GCCAAGCAGCAGACCTGAAACCTGCAGGACTTACACAAAGATGCAGAGATAGAACA 253
Qy 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 254 GCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTTGGAAC 313
Qy 61 IleIleSerAsnIleSerTyrrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
Db 314 ATCATTAGCAATATCAGTTACACAGCCACCACCTGCGGACGCTGACCAGCAATCTAAAT 373
Qy 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 374 GAAGTCAGACCCACTTGACACAGATACCTT-ACCAACACACACAGATGATCTGACCTCTTG 432
Qy 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 433 AATAATACCTGGCCACACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAGATTG 492
Qy 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
Db 493 ATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCATGATTATGAGAAATGAAG 552

```

Qy	141	LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro	160
Db	553	CTAGTACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAAGTACTACAAGGTCCACCG	612
Qy	161	GlyProAsrGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsn	180
Db	613	GGCCCCAGGGGTCCAAGAGTCAAGAGATCCCAAGGACCCCTGGCCCACTGGCAAC	672
Qy	181	LysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGluAArgGly	200
Db	673	AAGGACAGAAGAGAGAGAGGGGAGCCTGGACCACCTGGCCCTCGGGTGAAGAGGC	732
Qy	201	ProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln	220
Db	733	CCAATTGGACAGCTGCTCCCGGAGAGCGTGGCGCAAGAGATCTAAGAGCTCCACG	792
Qy	221	GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp	240
Db	793	GGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCGCAGTGGGAC	852
Qy	241	ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProProGly	260
Db	853	CCAGGCCCCCGGCCACACAGCAAGAGGACCTCCCGGCCCTCAGGGCCCTCTCTGGC	912
Qy	261	PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProAArgGlyLeuPro	280
Db	913	TTCCAGGACTTCAGGACCCGTTGGGAGCGTGGGTGCCTGGACCTCGGGAGTGCCA	972
Qy	281	GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro	300
Db	973	GGCTTGCTGGGTACAGGCATGCGAGGCCCAAGGCCCGCCCGGCCCTCTCTGGCCCA	1032
Qy	301	SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn	320
Db	1033	TCAGGAGCGGTGTGCCCTGGCCCTGCAGATGAGCCCAACCCCGGACCGGAGGACAA	1092
Qy	321	GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys	340
Db	1093	AGCTGCCCGCCTCAGTGGGAAGACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAA	1152
Qy	341	GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle	360
Db	1153	GAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTGTGTTTCATA	1212
Qy	361	AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyAArgGluSerHisTrp	380
Db	1213	ACACATAGAGAGGACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGACCACTGG	1272
Qy	381	IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro	400
Db	1273	ATCGGCTCACAGACTCAGAGCGTCAAAATGATGGAAGTGGCTGGATGGACATCTCCA	1332
Qy	401	AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly	420
Db	1333	GACTACAAAAATTTGGAAGCTGGACAGCGCGGATAACTGGGTCTATGGCATGGGCCAGGA	1392
Qy	421	GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal	440
Db	1393	GAAGACTGCTGGGTGTATTTATCTCGGGCAGTGGAAACGATTTCCTCAATGTGAAGAGTC	1452
Qy	441	AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457	
Db	1453	AATTAATCTTCAATTCGAAAAACAGAGGAGACAGTACTGTCATCTGCATTA 1503	
RESULT 10			
ABL88085			
ID ABL88085 standard; cdna; 2005 BP.			
XX			
AC ABL88085;			
XX			
DT 16-MAY-2002 (first entry)			
XX			

DE Human PRO7223 cDNA sequence SEQ ID NO:27.

XX Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;
KW vulnery; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial stenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping; gene; ss.

XX Homo sapiens.

OS WO200200690-A2.

XX 03-JAN-2002.

XX 20-JUN-2001; 201WO-US19692.

XX 23-JUN-2000; 2000US-213637P.

XX 20-JUL-2000; 2000US-219556P.

PR 25-JUL-2000; 2000US-220624P.

PR 25-JUL-2000; 2000US-220664P.

PR 28-JUL-2000; 2000WO-US20710.

PR 02-AUG-2000; 2000US-222695P.

PR 17-AUG-2000; 2000US-0643657.

PR 23-AUG-2000; 2000WO-US23522.

PR 24-AUG-2000; 2000WO-US23328.

PR 07-SEP-2000; 2000US-230978P.

PR 18-SEP-2000; 2000US-0664610.

PR 18-SEP-2000; 2000US-0665320.

PR 24-OCT-2000; 2000US-242925P.

PR 08-NOV-2000; 2000US-0709238.

PR 08-NOV-2000; 2000WO-US30952.

PR 10-NOV-2000; 2000WO-US30873.

PR 01-DEC-2000; 2000WO-US32678.

PR 20-DEC-2000; 2000US-0747259.

PR 20-DEC-2000; 2000WO-US34956.

PR 22-JAN-2001; 2001US-0767609.

PR 28-FEB-2001; 2001US-0796498.

PR 28-FEB-2001; 2001WO-US06520.

PR 01-MAR-2001; 2001WO-US06666.

PR 09-MAR-2001; 2001US-0802706.

PR 14-MAR-2001; 2001US-0808689.

PR 22-MAR-2001; 2001US-0816744.

PR 05-APR-2001; 2001US-0828366.

PR 10-MAY-2001; 2001US-0854208.

PR 10-MAY-2001; 2001US-0854280.

PR 25-MAY-2001; 2001US-0866028.

PR 25-MAY-2001; 2001US-0866034.

PR 25-MAY-2001; 2001WO-US17092.

PR 30-MAY-2001; 2001US-0870574.

PR 30-MAY-2001; 2001WO-US17443.

PR 01-JUN-2001; 2001WO-US17800.

XX (GETH) GENENTECH INC.

XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NP;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX WPI: 2002-090516/12.

DR P-PSDB; ABB84830.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX Claim 2; Fig 27; 565pp; English.

PS ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
XX ABB85003. The PRO proteins and polynucleotides have cardiant, cytotatic
CC antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic
CC

CC activities, and can be used in gene therapy. The PRO polynucleotides,
 CC proteins, agonists and antagonists are useful for treating or diagnosing
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
 CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
 CC degeneration, atherosclerosis, hypertension, arterial restenosis,
 CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
 CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
 CC carcinoma) and wound healing. The PRO polynucleotides have applications
 CC in molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
 CC probes used in the exemplification of the present invention.

XX
 SQ Sequence 2005 BP: 606 A; 493 C; 491 G; 415 T; 0 other;

Alignment Scores:

Pred. No.: 2,48e-79 Length: 2005
 Score: 2487.00 Matches: 455
 Percent Similarity: 99.56% Conservative: 0
 Best Local Similarity: 99.56% Mismatches: 2
 Query Match: 98.69% Indels: 1
 DB: 24 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x ABL88085 (1-2005)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB 134 ATGAACAGCCAGCTCAACTTCACAGTCAGATGAGAGAACATCACCACCTATCTCTCAA 193
 QY 21 AlaAsnGluInAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 DB 194 GCCAACGAGCAAGCCTGAAGACCTGCGAGGACTTACACAAAGATGCGAGAGAAATAGACA 253
 QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 DB 254 GCCATCAAGTTCAACCACTGGAGGACGCTCCAGCTCTTGAGACGGATTTGTGAAC 313
 QY 61 IleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 DB 314 ATCATTTAGCAATATCAGTTACAGCCACCCACCTGCGAGCGCTGACCCAGCAATCTAAAT 373
 QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSerLeu 100
 DB 374 GAAGTCAGACCACTTGACAGATACCTT-ACCAACACACAGATGATCTGACCTCCCTTG 432
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 DB 433 AATAATACCTGGCCCAACATCCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 492
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 DB 493 ATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG 552
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 DB 553 CTAGTAGACTCCACGATGCTGAGTCATCAAGAATTTTACAATACTACAAGTCCACCG 612
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
 DB 613 GSCCCAGGGTCCAAAGAGTCCAGAGATCCAGGAGACCCCTGGCCCACTGGCAAC 672
 QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
 DB 673 AAGGACAGAAAGAGAGAAAGGGGAGCGCTGGACCACTGGCCCTGCGGGTCAGAGAGGC 732
 QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 DB 733 CCATTTGACCACTGGTCCCGGAGAGCGTGGCGCAAGAGATCTAAAGGCTCCCGAG 792
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 DB 793 GSCCCCAAGGCTCCCGTGGTTCCCTGGGAGAGCCCGGCTCAGGGCCCACTGGGAC 852
 QY 241 ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProGly 260

DB 853 CCAGGCCCCCGGGCCCGCCAGGCAAGAGGAGACTCCCGGGCCCTCAGGGCCCTCTCTGCG 912
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 DB 913 TTCCAGGAGACTTCAGGGCACCCTTGGGAGCCTGGGGTCCCTGGAGCTCGGGACTGCCA 972
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 DB 973 GGCTTGGCTGGGTACCGAGCATGCCAGGCCCAAGGGCCCCCGGGCCCTCTCTGCCCCA 1032
 QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 DB 1033 TCAGGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCAACCCCGGACCGGAGGACAAT 1092
 QY 321 GlyCysProProHisThrLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 DB 1093 AGCTGCCCGCCCTACTGGGAAGACTTCACAGACAATGCTACTATTTTCAGTTGAGAAA 1152
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 DB 1153 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGCTTTCACATCTTGTTCATA 1212
 QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 DB 1213 AACACTAGAGAGAACAGCAATGGATAAAAAAACAGATGCTAGGAGAGAGAGCCACTGG 1272
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 DB 1273 ATCGGCCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGTGGATGGGACATCTCCA 1332
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 DB 1333 GACTACAAAAATTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGCCAGGA 1392
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 DB 1393 GAAGACTGCTGCTGGGTGATTATGCTGGGCGAGTGGAGCATTTCCAATGTGAAGACGTC 1452
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 DB 1453 AATAACTTCATTGGCAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 1503

RESULT 11
 AAH43037
 ID AAH43037 standard; cDNA; 2637 BP.
 XX
 AC AAH43037;
 XX
 DT 15-OCT-2001 (first entry)
 XX
 DE Nucleotide sequence of a human scavenger receptor.
 XX
 KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 92..2320
 FT /*tag= a
 FT /product= "scavenger receptor"
 XX
 PN WO200159107-A1.
 XX
 PD 16-AUG-2001.
 XX
 PF 08-FEB-2001; 2001WO-JP00874.
 XX
 PR 14-FEB-2000; 2000JP-0035155.
 PR 10-OCT-2000; 2000JP-0309068.
 XX

/product= "scavenger receptor"

FT
XX WO200159107-A1.
XX
XX PD 16-AUG-2001.
XX
XX PF 08-FEB-2001; 2001WO-JP00874.
XX
XX PR 14-FEB-2000; 2000JP-0035155.
XX
XX PR 10-OCT-2000; 2000JP-0309068.
XX
XX PA (FUSO) FUSO PHARM IND LTD.
XX
XX PI Wakamiya N;
XX
XX DR WPI; 2001-497076/54.
XX
XX DR P-PSDB; AAG63350.
XX
XX PT New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation -
XX
XX PS Claim 4; Page 105-109; 118pp; Japanese.
XX
XX CC The present sequence encodes a human scavenger receptor, designated
XX SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Alignment Scores:
Pred. No.: 4.19e-53 Length: 2256
Score: 1738.00 Matches: 333
Percent Similarity: 72.87% Conservative: 0
Best Local Similarity: 72.87% Mismatches: 0
Query Match: 68.97% Indels: 124
DB: 22 Gaps: 1

US-09-763-712a-2_COPY_91_547 (1-457) x AAH43054 (1-2256)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
DB 929 ATGAACAGCCAGCTCAACTCATTCACAGGTCAGATGGAGAACATCACCCTATCTCTCAA 988

QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
DB 989 GCCAAGGAGCAACCTGAAGACCTGCAGGACTTACACAAGATGCAGAGATAGAAC 1048

QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
DB 1049 GCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTGGAGCGGATATTGTGAAC 1108

QY 61 IleIleSerAsnIleSerThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
DB 1109 ATCATAGCAATATCAGTTACACAGCCACACCTCGCGAGCCTGACCAAGCAATCTAAAT 1168

QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSerLeu 100
DB 1169 GAAGTCAGGACCACCTGGACACATACCTTACCAACACACACAGATGATCTGACCTCCCTTG 1228

QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
DB 1229 AATAATACCTGGCCCAACATCGTTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 1288

QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
DB 1289 ATGAGGTGAGGTTAGACACTGAAGTAGCAACTTATCAGTCATTATGGAAGAAATGAAG 1348

QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
DB 1349 CTAGTAGACTCCAGCATGCTCAGCTCATCAAGAATTTTACATATACTACAAGTCCACCG 1408

QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
DB 1409 GGGCCAGGGTCCAAAGAGGTGACAGAGATCCAGGAGCCCTGGCCCACTGGCAAC 1468

QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGluGluArgGly 200
DB 1469 AAGGACAGAAAGAGAGAGGGGAGCTGGACCACCTGGCCCTGCG - 1516

QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
DB 1516 - 1516

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
DB 1516 - 1516

QY 241 ProGlyProProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProProGly 260
DB 1516 - 1516

QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
DB 1516 - 1516

QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
DB 1516 - 1516

QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
DB 1516 - 1516

QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
DB 1517 GGCTGCCCGCCCTCACTGGGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAA 1576

QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
DB 1577 GAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAACTTTCACATCTTGTTCATATA 1636

QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
DB 1637 AACACTAGAGAGAAACAGCAATGGATAAAAAACAGATGTAGGGAGAGAGCCACTGG 1696

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
DB 1697 ATCGGCTCACAGACTCACAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1756

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
DB 1757 GACTACAAAATTTGGAAGTGGACAGCGGATACTGGGGTTCATGGCCATGGGCCAGGA 1816

QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
DB 1817 GAAGACTGTGTGGGTGATTTATGCTGGGCAGTGAAGCATTTCCAATGTGAAGACGTC 1876

QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
DB 1877 ATACTTCATTTTGGAAAAACAGAGGAGAGAGACTGCTCATCTGCATTA 1927

RESULT 13
AAS71133
ID AAS71133 standard; cDNA; 1521 BP.
XX
XX AAS71133;
XX
XX 13-FEB-2002 (first entry)
XX

DE DNA encoding novel human diagnostic protein #6937.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG06946.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID NO 6937; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AA564197-AA594564 represent novel human
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1521 BP; 386 A; 420 C; 425 G; 290 T; 0 other;

Alignment Scores:

Pred. No.: 1.92e-52 Length: 1521
 Score: 1715.50 Matches: 298
 Percent Similarity: 94.43% Conservative: 7
 Best Local Similarity: 92.26% Mismatches: 7
 Query Match: 68.08% Indels: 11
 DB: 23 Gaps: 1

US-09-763-712A-2_COPY_91_547 (1-457) x AAS71133 (1-1521)

QY 132 LeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLys 151

Db 130 CTACATGTCTCTATTATGCAATGAGAAATCTACCCACAGG----- 171

QY 152 AsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer 171

Db 172 -----CAAGGTCCACCGGGCCCGAGGGTCCRAAGAGTGCACAGAGATCC 216

QY 172 GlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlyGluProGly 191

XX

16-11-03

Db	217		CAGGGACCCCTGTGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGAGCCTGGA	276
QY	192		ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg	211
Db	277		CCACCTGGCCCTGCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCCCGGAGACGT	336
QY	212		GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys	231
Db	337		GGCGCAAGAGGATCTAAAGGCTCCAGGGCCCCCAAGGCTCCCGTGTTCCTCGGAAG	396
QY	232		ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGly	251
Db	397		CCCGGCCCTCAGGGCTCCAGTGGGACCCAGGCCCGCCCGGCCACAGCAAGAGGGA	456
QY	252		LeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluPro	271
Db	457		CTCCCCGGGCCCTCAGGGCCCTCTGTGGCTTCCAGGGACTTCAGGGGACCGTTGGGAGCCT	516
QY	272		GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro	291
Db	517		GGGGTGCCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGCC	576
QY	292		LysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsn	311
Db	577		AAGGGCCCCCGGCCCTCTGTGCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAA	636
QY	312		GluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAsp	331
Db	637		GAGCCAAACCCAGCACCGGAGGACAATGGTGGCCGCCCTCAGTGGGAAGAACTTCACAG	696
QY	332		LysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGlu	351
Db	697		AAATGCTACTATTTTCACTTGAGAAAGAAATTTTGGAGTGCAGAAAGCTTTTCTGTGA	756
QY	352		AspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLys	371
Db	757		GACAATCTCTTCATCTTTTTCATAAACACTTAGAGAGGAAACAGCAATGGATAAAAA	816
QY	372		GlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGlu	391
Db	817		CAGATGGTAGGAGAGAGAGAGCCACTGGATCGGCCCTCACAGCTCAGAGCGTGAAATGAA	876
QY	392		TrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAsp	411
Db	877		TGSAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGAT	936
QY	412		AsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGln	431
Db	937		AACTGGGGTCATGGCCATGGGCCAGGAGAGACTGTGCTGGGTGTGATTTATGCTGGGCAG	996
QY	432		TrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThr	451
Db	997		TGGAAGCATTTCCATGTGAAGACGTCATACTTCATTTGCGAAAAAGACAGGGAGACA	1056
QY	452		ValLeuSer 454	
Db	1057		GTCTTCACG 1065	
RESULT 14				
AAAX27858				
ID	AAAX27858	standard; DNA; 3685 BP.		
XX	AAAX27858;			
AC	AAAX27858;			
XX	02-JUN-1999	(first entry)		
DT	02-JUN-1999	(first entry)		
DE	Human CSR3	protein coding sequence.		
XX				
KW		Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;		
KW		scavenger receptor protein; intracellular stress; arteriosclerosis;		
KW		diabetic circulatory obstruction; microbial infection; ss.		
XX				

Search completed: March 21, 2003, 08:38:02
Job time : 262.038 secs

Pred. No.:	5.17e-16	Length:	3810
Score:	682.00	Matches:	132
Percent Similarity:	62.42%	Conservative:	54
Best Local Similarity:	44.30%	Mismatches:	94
Query Match:	27.06%	Indels:	18
DB:	20	Gaps:	2

Qy 11 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 30

Db 1203 CAGCTGGATAACATCTCGTCTTCTCTGGATGACCACGAGAGAACATGCATGATCTTCAG 1262

Db 1263 TACCATACCCACTACGCCCCAGAACCGCACTGTGGAGAGGTTTGAGTCTCTGGAGGACGC 1322

1223

QY /I H1sLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 90
|||:: ::::: ||:::||||| :::|||||

Qy 91 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 110

QY 111 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 130

```

131  Asn[enSerVal]leMetGlcC"MetValLeuVal"AsnSerIleValGlcVal"leIleVal" 150

```

UD 1563 AACCTCTCCATGATCGTGGAGSAGATGAAGGCAGTGGACACACAGCATGGAGAAATCCTT 1622

Db 1623 CGCAATGTCACCATCTCTACGAGGTGCCCCCGGCCTCCAGGACCAAGGATTCAAAGGA 1682

Db 1683 GATATGGCGGTGAAGGGCCCTGTTGGCGGCAAGAGCCCCGAAAGGACACCCCCGGCATCTTG 1742

DB 1743 GGGCCCTTGGGACCCCAAGGCTCCTCAGCGGCACTCGACAGCGGCGCCCTCTGGCGACACA 1802

QY 199 ARGGLYPTOLECTYPROALAGLYPROPTROGLYGLNARGGLYGLYLYSGLYSERLYSGLY 218

Qy 219 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 238
... ..
... ..

QY 239 GlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyPro 258

[illegible]

Db 1965 CCGGGAAATTGCAGGGAAGACACAGGGTCACACAGGCCAGCGGGGGCCATGGGGCCCTAAGGGT 2024

Db 2025 GAACGAGGATCCAGGGTCCCCCTGGTCTCCCGGGGCCTCCAGGTCCACCAGGA 2078

24

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:26:21 ; Search time 48.552 seconds
(without alignments)
2886.622 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547

Perfect score: 2520

Sequence: 1 MNSQNSPTGOMENTITISQ.....EDVNNFICKDRETVLSSAL 457

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09763712/runat_14032003_100951_18187/app_query.fasta_1.1877
-DB=Issued_Patents.NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn.1.108 @runat_14032003_100951_18187 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents.NA.*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	459.5	18.2	810	1	US-08-642-255-60
2	452.5	18.0	3181	1	US-08-655-086-1
3	448.5	17.8	756	1	US-08-642-255-50
4	436.5	17.3	1608	4	US-09-029-348-19
5	428	17.0	4031	1	US-08-159-784-1
6	424.5	16.8	1572	4	US-09-297-269-39
7	419	16.6	2543	1	US-08-555-669-11
8	419	16.6	2543	3	US-09-073-663-11
9	412.5	16.4	432	1	US-08-642-255-48
10	409	16.2	3394	1	US-08-159-784-4
11	402.5	16.0	1416	1	US-07-621-091G-1
12	402.5	16.0	1416	2	US-07-399-889-1

13	402.5	16.0	1416	3	US-09-167-364-1	Sequence 1, Appli
14	402.5	16.0	1416	4	US-09-439-897-1	Sequence 1, Appli
15	400.5	15.9	4359	4	US-09-484-970B-4	Sequence 4, Appli
16	399	15.8	1881	4	US-09-029-348-20	Sequence 20, Appli
17	394.5	15.7	1560	2	US-08-794-795-5	Sequence 5, Appli
18	394.5	15.7	1560	4	US-09-249-200-5	Sequence 5, Appli
19	394	15.6	5102	1	US-08-494-168-1	Sequence 1, Appli
20	393.5	15.6	503	4	US-09-297-269-40	Sequence 40, Appli
21	391	15.5	1560	4	US-09-453-702B-264	Sequence 264, App
22	391	15.5	61663	4	US-09-453-702B-62	Sequence 62, Appli
23	389.5	15.5	1703	2	US-08-794-795-1	Sequence 1, Appli
24	389.5	15.5	1703	4	US-09-249-200-1	Sequence 1, Appli
25	380.5	15.1	1868	1	US-08-392-367B-1	Sequence 1, Appli
26	380.5	15.1	1868	3	US-08-893-467A-1	Sequence 1, Appli
27	356	14.1	1839	1	US-08-383-744-1	Sequence 1, Appli
28	356	14.1	1839	2	US-08-999-336-1	Sequence 1, Appli
29	356	14.1	1839	5	PCT-US96-01427-1	Sequence 1, Appli
30	355.5	14.1	2383	4	US-09-523-487-9	Sequence 9, Appli
31	355.5	14.1	2409	3	US-09-320-095-9	Sequence 9, Appli
32	350	13.9	1458	4	US-09-111-470-3	Sequence 3, Appli
33	343	13.6	506	1	US-08-253-155A-61	Sequence 61, Appli
34	331	13.1	2824	2	US-09-010-928B-3	Sequence 3, Appli
35	330.5	13.1	392	4	US-09-404-879A-372	Sequence 372, App
36	326.5	13.0	48908	4	US-09-453-702B-137	Sequence 137, App
37	325.5	12.9	9827	4	US-09-453-702B-66	Sequence 66, Appli
38	323.5	12.8	3404	4	US-09-453-702B-94	Sequence 94, Appli
39	322	12.8	1370	4	US-09-111-470-9	Sequence 9, Appli
40	320.5	12.7	45175	4	US-09-453-702B-116	Sequence 116, App
41	315.5	12.5	2851	4	US-09-535-521-1	Sequence 1, Appli
42	315.5	12.5	2851	4	US-09-535-521-3	Sequence 3, Appli
43	310.5	12.3	876	4	US-09-535-521-4	Sequence 4, Appli
44	310.5	12.3	876	4	US-09-535-521-6	Sequence 6, Appli
45	306.5	12.2	38584	4	US-09-453-702B-50	Sequence 50, Appli

ALIGNMENTS

RESULT 1

US-08-642-255-60
; Sequence 60, Application US/086422255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

```

:
: COUNTRY: US
: ZIP: 11553
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0. Version #1.25
:

```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/655,086
3 FILING DATE: 03-JUN-1996
4 CLASSIFICATION: 435
5 ATTORNEY/AGENT INFORMATION:
6 NAME: STEEN, JEFFREY S.
7 REGISTRATION NUMBER: 32,063
8 REFERENCE/DOCKET NUMBER: 203-1632

```

TELECOMMUNICATION INFORMATION;
TELEPHONE: 516-228-8484

```

/ BLASTX: J16248.6310
/
/ INFORMATION FOR SEQ ID NO: 1:
/
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 3181 base pairs
/   TYPE: nucleic acid
/   STRANDEDNESS: single
/   TOPOLOGY: unknown
/
/ MOLECULE TYPE: cDNA
/

```

0.5 0.0 0.55 0.60 1

Alignment Scores:

Pred. No.: 3.58e-17 Length: 3181

Score:	452.50	Matches:	33
Percent Similarity:	52.11%	Conservative:	12
Best Local Similarity:	46.48%	Mismatches:	53
Query Match:	17.96%	Indels:	49

US-09-763-712A-2_COPY_91_547 (1-457) x US-08-655-086-1 (1-3181)

QY 156 leuGlnuLyPFIoPFIoAGLyAspArgGlySeiGlnuLyPFIoF 173
:::||||| ||||| ||| |||||::: ||| |||
Db 893 GTTCAGGACCCCTGGCCCTGCTGGAGGAGGAAGCGAGGAGCTCGAGGTGAACCC 952

QY 1/6 GRYPRINRGVASNLGSGYGNNLYSGYGYNNLYSGYGRUPLD..... 190
||||| ||||| ||| |||||::||| |||
Db 953 GGACCCACTGGCTGCCCGGACCCTTGGCAGCGTGTTGGACCTGGTAGCGGTTC 1012

Qy	191	-----GlyProPogLyProAlaGlyGuArgGlyProIleGly	203
Db	1013	CCTGGCCAGATGGTGTGCTGGTCCCACAGGGTCCCGCTGGTGACGTGGTCTCCTGGC	1072

[illegible]

QY	221	GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp	240
		:::	
Db	1133	GGTGCCAAAGGGTCTGACTGGAGGCCCTGGACGCCCTGTCTGTATGGCAAACT	1186

Qy	241	ProGlyProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProPro---	259
Db	1187	---GGCCCCCCTGTCCCGCCGGTCAAGATGTCGCCCCCGACCCCAGGCCACCCTGGT	1243

Qy	260	-----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro	271
		::	
Db	1244	GCCCGTGGTCAGGCTGGTGTATGGGATTTCCCTGGACCTAAAGGTGCTGCTGGAGAGCCC	1303

Qy	272	GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal	-----	285
		::		
Db	1304	GGCAAGGCTGGAGAGCGAGGTGTCCCGACCCCTGGCGCTGTGGTCTCGTGGCAAA		1363

Qy	286	ProGlyMetProGlyProLysGlyProGlyProGlyProGlyProSerGly-----	302
		:::::::::::::: :::	
Db	1364	GATGGAGAGGCTGGAGCTCAGGAGACCCCTGGCCCTGCTGTCCTGGGAGAGAGGT	1423

QY 303 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 316
 Db 1424 GAACAAGGCCCTGCTGGTCCCTCC-----GGATCCAGGGTCTCCTGGTCTCTGCTGCTCT 1480
 QY 317 -----ProGluAspAsnGlyCysPro 323
 Db 1481 CCAGGTGAAGCAGGCAACCTGGTGAACAGGGTGTCTCT 1519
 RESULT 3
 US-08-642-255-50
 ; Sequence 50, Application US/08642255
 ; Patent No. 5773249
 ; GENERAL INFORMATION:
 ; APPLICANT: CAPPELLO, Joseph
 ; APPLICANT: FERRARI, Franco A.
 ; TITLE OF INVENTION: High Molecular Weight Collagen-Like
 ; TITLE OF INVENTION: Protein Polymers
 ; NUMBER OF SEQUENCES: 135
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/642,255
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROWLAND, Bertram I.
 ; REGISTRATION NUMBER: 20,015
 ; REFERENCE/DOCKET NUMBER: A5556-3/BI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 494-8700
 ; TELEFAX: (415) 494-8771
 ; TELEX: 910 277299 FHT UR
 ; INFORMATION FOR SEQ ID NO: 50:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 756 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: other nucleic acid
 ; DESCRIPTION: /desc = "synthetic"
 US-08-642-255-50
 Alignment Scores:
 Pred. No.: 1.35e-17 Length: 756
 Score: 448.50 Matches: 99
 Percent Similarity: 52.70% Conservative: 18
 Best Local Similarity: 44.59% Mismatches: 85
 Query Match: 17.80% Indels: 21
 DB: 1 Gaps: 5
 US-09-763-712a-2_COPY_91_547 (1-457) x US-08-642-255-50 (1-756)
 QY 107 IleArgLeuAspSerValSerLeuArgMetGlnAspLeuMet-----ArgSerArg 124
 Db 56 GTGGCGCTGGACGGGTGTCTCCACCGGCTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGC 115
 QY 125 LeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer 144
 Db 116 CTGGACCGGCTGTCTCCACCGG-----GTGCTCCGGGACCTGCAGGCCCGCCAGGTG 166
 QY 145 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGlyProArgGly 164
 Db 167 CGCCTGGACCGGCTGTCTCCACCGG-----GTGCTCCG--GGACCTGCAGGCCCGCCAGGT 219

QY 165 ProArgGlyAspArgSerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLys 184
 Db 220 GGCCTGGACCGGCTGTCTCCACCGGCTGTCTCCGGACCTGCAGGCCCGCCAGGTGCGCCT 279
 QY 185 GlyGluLysGlyGluProProGlyProProGlyProAlaGlyGluArgGlyProIleGlyPro 204
 Db 280 GGACCGGCTGTCTCCACCGGCTGTCTCCGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCG 339
 QY 205 AlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGly 224
 Db 340 GCTGTGTCCACCGGCTGTCTCCGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGT 399
 QY 225 SerArgGlySerPro-----GlyLysProGlnGlyProGlnGlyProSerGlyAspPro 241
 Db 400 CCACCGGCTGTCTCCGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTCTCCACCG 459
 QY 242 GlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGlyPhe 261
 Db 460 GGTGTCTCCGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTCTCCACCGGCTGT 519
 QY 262 GlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGly 281
 Db 520 CCGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTCTCCACCGGCTGTCTCCGGGA 579
 QY 282 LeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyProSer 301
 Db 580 CTTGCAGGCCCGCCAGGTGCGCCTGCAGCGGCTGTCTCCACCGGCTGTCTCCGGACCTGCA 639
 QY 302 GlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 321
 Db 640 GGC-----CCGCCAGGTGCGCCTGGACCGGCTGTCTCCACCGGCTGTCT 669
 QY 322 CysPro 323
 Db 670 CCACCG 675
 RESULT 4
 US-09-029-348-19
 ; Sequence 19, Application US/09029348
 ; Patent No. 6171827
 ; GENERAL INFORMATION:
 ; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
 ; TITLE OF INVENTION: NOVEL PROCOLLAGENS
 ; FILE REFERENCE: G087857PUS LISTING
 ; CURRENT APPLICATION NUMBER: US/09/029,348
 ; CURRENT FILING DATE: 1998-05-07
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 19
 ; LENGTH: 1608
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
 ; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
 US-09-029-348-19
 Alignment Scores:
 Pred. No.: 1.38e-16 Length: 1608
 Score: 436.50 Matches: 102
 Percent Similarity: 47.06% Conservative: 18
 Best Local Similarity: 40.00% Mismatches: 67
 Query Match: 17.32% Indels: 70
 DB: 4 Gaps: 9
 US-09-763-712a-2_COPY_91_547 (1-457) x US-09-029-348-19 (1-1608)
 QY 145 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGlyProArgGly 164
 Db 59 CAACATGCCCACTTTACAGAGGAA--ACTGTAAAGAAAGGCCCGCCAGGTAGAGGA 117
 QY 165 ProArgGlyAspArgSerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLys 184

Db 118 CCACGTGGAGAGGGTCCACAGGCCCCCAGGACAGATGGTGAAGATGTCACCA 177
QY 185 GlyGluLysGlyGluProGlyProGlyPro----- 195
Db 178 GGCCCTCCTGGTCCACCTGGTCTCTCTGCCCCCTGGTCTGGTGGGACATTGGTGT 237
QY 196 -----AlaGlyGluArgGlyProLeuProAla 205
Db 238 CAGTATGATGAAAGAGTGGACTTGGCCCTGGACCAATGGCTTAATGGACCTAGA 297
QY 206 GlyProGlyGluArgGlyGlyLysGlySerGlnGlyProLysGlySer 225
Db 298 GGCCCACTGGTGCAGCTGGAGCCCCAGGCCCTCAAGGTTTCCAAAGGACCTGCTGGTGA 357
QY 226 ArgGlySerProGlyLys-----ProGlyProGlnGlyPro----- 237
Db 358 CTGGTGAACCTGGTCAAACTGGTCTGTCAGGTGACCTGGTCTCATGCCCCCTGGGT 417
QY 238 -----SerGlyAspProGlyProGlyProGlyProGlyLysGlu 250
Db 418 CTGCTGGCAACATGAAACCGTGTGAAACTGGTCTCTGCTCTGTTGGTCTGCT 477
QY 251 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 270
Db 478 GGTGCTGTGGCCCAAGAGGTCTAGTGGCCCCACAAAGGATTCGTGGCGATAAGGAGAG 537
QY 271 ProGlyValProGlyProArgGlyLeuPro-----Gly 281
Db 538 CCGGGTGAAGAGGGGCCAGAGGTCTCTCTGCTTCAAGGACACAAATGGATTGCAAGGT 597
QY 282 LeuProGlyVal-----ProGlyMetProGlyProLys 292
Db 598 CTGCTGGTATCGTCTGCTCACCATGTGTCAAGGTGCTCTGCTGCTCGCTGGTCTGCT 657
QY 293 GlyProProGlyProGlyProGlyProSerGly----- 302
Db 658 GGTCTTAGGGGCCCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
QY 303 ---AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 321
Db 718 CTGGTACGGTGGACCTGCTGG-CATTGAGGCCCTCAGGTGCTCACCAGGCC----- 770
QY 322 Cys---ProPro-----HisTrpLysAsnPheThrAspLysCys 333
Db 771 TGTGGCCCCCTGGTCCCTGGCCCTCCTGGACCTCCTGACCTCAGGTGT 815

RESULT 5

US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4031
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-159-784-1
Alignment Scores:
Pred. No.: 1.07e-15 Length: 4031
Score: 428.00 Matches: 95
Percent Similarity: 46.41% Conservative: 15
Best Local Similarity: 40.08% Mismatches: 49
Query Match: 16.98% Indels: 78
DB: 1 Gaps: 8
US-09-763-712a-2_COPY_91_547 (1-457) x US-08-159-784-1 (1-4031)
QY 141 LeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProPro 160
Db 976 GTGGTCCAGAGCCCCAACCTCAACAACTGTCTCCCTGGA-----GCACAAGAGCCCCCG 1026
QY 161 GlyProArgGlyPro----- 165
Db 1027 GGACCTCAGGGGCCACCAGGAAGGATGTCACCTCCAGGAAGGATGGTGAACGGGTGAC 1086
QY 166 -----ArgGlyAspArgGlySerGlnGlyProProGlyProThrGly 179
Db 1087 CCTGGTGAAGATGGGAGACCGGGTGTACACTGGACCTCAAGGCTTTCAGGGAGCCCCAGGA 1146
QY 180 AsnLysGlyGlnLysGlyGluLysGlyGluPro-----GlyProProGly 194
Db 1147 GATGGGCCCTCAAGGGCGAGGAGATCCTGTATTGGGCCCGCCGAGACCTCCAGGG 1206
QY 195 ProAlaGlyGluArgGlyProLeu----- 202
Db 1207 CCTCCAGGGCCACCAGGACCTCTCTTCAGACAAGACAAGCTGACCTTCATTGACATGGAG 1266
QY 203 -----GlyPro 204
Db 1267 GGATCCGGTTTCAGCGGAGACATAGAGAGCCTTAGAGGCCACAGAGGCTTCCCTGGCCCC 1326
QY 205 AlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGly 218
Db 1327 CCGGGCCCCCTGGTGTCTCCAGGACTTCTGTGTGTCAGGAGGAGGCTTGGGATCAATGGT 1386
QY 219 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 238
Db 1387 TCCATGACCA--GGACCTGCAGGCCCTCTCTGGTGTACCTGGGAAGGAGGCCCCCC 1443
QY 239 GlyAspProGlyProGlyProGlyLysGlyGluLeuProGlyProGlnGlyPro 258
Db 1444 GGTTCCTCAGGTCCCCCGGACCTCCAGGTCT-----CCAGGCAAGAGGCCCA 1494
QY 259 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg--- 277
Db 1495 CCAGGAGTGGCGCCAGAAAGGAGGTGTGTGTGTGGTCATCCAGGACCCCAAGGGG 1554
QY 278 -----GlyLeuProGlyLeuProGlyValProGlyProGlyMetPro 289
Db 1555 AGCAAGAGGACCTTGGGCCCATCGGTATCGCTGGCAAGCTGTGGCTGTGGTGTGATCCCT 1614
QY 290 GlyProLysGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 306


```
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-08-555-669-11

Alignment Scores:
Pred. No.: 2,11e-15      Length: 2543
Score: 419.00           Matches: 121
Percent Similarity: 35.59% Conservative: 26
Best Local Similarity: 29.30% Mismatches: 91
Query Match: 16.63%      Indels: 176
DB: 1                    Gaps: 9

US-09-763-712a-2_COPY_91_547 (1-457) x US-08-555-669-11 (1-2543)
QY 142 ValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGly 161
DB 986 CTCGATGGCCAGGAAGGAGAGGCTGGTCCGAAC-----GGTGGCTCGGGA 1030
QY 162 ProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys 181
DB 1031 GAGAAGGGCCCCAACGGGCTCCCGGCTCCCTGGAGAGCGGGGTCCTCAAGGCGAGAAG 1090
QY 182 GlyGlnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGlyPro 201
DB 1091 GGAGAACGGGCGAGAGCTGGGGAGCTGGGTGAGGCGCGGCCCTCTGGAGAGCCAGCGCTC 1150
QY 202 IleGlyProAlaGlyProProGlyGluArgGlyGlyGlySerLysGlySer----- 219
DB 1151 CTGGAGATGCTGGCATGCTGGGGAGCGCGGTGAGGCTGGCCACCGGGCTCAGCGGGG 1210
QY 220 -----GlnGlyProLys 223
DB 1211 GCCTCGGCCCAACAGCCCTCCCGGAGCCCTGGTGTCCGAGGCTTCAGGCGCCAGAAG 1270
QY 224 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 243
DB 1271 GGCAGCATGGGAGACCCCGGCTCTCCAGGCCCCAGGGCTCCGAGGTGACGTGGGCGAC 1330
QY 244 ProGlyProGlyLysGlnGlyLeuProGlyProGlnGlyProProGlyPheGlnGly 263
DB 1331 CGGGTCCGGGAGTGCCCAAGGCCCTAAAGGAGACCAAGGTATTGACGGTTCGACGGT 1390
QY 264 LeuGlnGlyThrValGlnGly----- 270
DB 1391 CTTCCTGGGATAAAGAGAGACTGGTCCAGCGGCTGGTCGACCCCAAGAGAGAGTCT 1450
QY 270 ----- 270
DB 1451 GGCAGTCGAGGGAGAGTGGGCCCCCAAGGACCCAGGGTCCCAACGGCACCGCGTGT 1510
QY 271 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 287
DB 1511 CAGGGTGTCCCGGGCCCCCGGCTCTCTGGGCTCGAGGGGCTCCCGGGTGTCTCTGGC 1570
QY 288 Met----- 288
DB 1571 ATCAGCGGGAAGCGGGAGTTCCGGGGAAGGAGGCCAGCGAGCGCATCAGGGAGCTG 1630
QY 288 ----- 288
DB 1631 TGTGGGGGATGATCAGCAACAAATTCACAGTTAGCGCGCACCTAAGGAAGCCCTTTG 1690
QY 289 -----ProGlyProLysGlyProProGlyProProGlyProSer 301
DB 1691 GCACCGGGTCCATTGGTCGGCGCGGTCCAGGTGGGCCCCCTGGGCCCCCAGGACCCCA 1750
QY 302 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 313
DB 1751 GGCCTCATTTGGTCACCTTGGCGCTCGAGGACCCCGCGGATACCGCGGTCCCTCGGGAG 1810
```

```
QY 314 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs 331
DB 1811 CTGGGAGACCCCGGGCCAGAG----- 1832
QY 331 pLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG1 351
DB 1833 -----GAAACCAAGGTGCACAGAGAGACAAG----- 1859
QY 351 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLy 371
DB 1860 -----GCGCGGCGAGGAGCAG----- 1874
QY 371 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG1 391
DB 1875 -----GGCTGGACGGCTCAAGAGACACGAGGCGCCCAAGGA 1912
QY 391 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAs 411
DB 1913 CCCCAGGCG-CTCCCGGCGACCAAGCAAGGACGCGGCGGAGCGTCTCCCGCGAGCCTGG 1971
QY 411 pAsnTrpGlyHisGlyHisGlyProGly----- 420
DB 1972 GCCTCCCGGAGATCTCGGGCTTCAGGTGCCATTGGGGCCCGGACACCGGGGATCTG 2031
QY 421 -----GluAspCysAlaGlyLeuIleTyrAlaGly 430
DB 2032 CGACACCTCAGCCTGCCAAGGAGCGCGTGTAGGAGGG 2068

RESULT 8
US-09-073-663-11
; Sequence 11, Application US/09073663
; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,663
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; NAME/KEY: CDS
; LOCATION: 47..2098
; US-09-073-663-11
```

Qy	142	ValAspSerLysHisGlycInLeuIleLysAsnPheThrIleLeuGlnGlyProProGly	161
Db	986	CTCATGGCCAGGAAGGAGCTGGTCGCAAC-----GGTGCTCCGGGA	1030
Qy	162	ProArgGlyProArGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys	181
Db	1031	GAAGAGGCCCAACGGCTGCCGGGCTCTCTGACAGCGGGGTCCAAGGCGAGAAG	1090
Qy	182	GlyGlnLysGlycInLysGlyGluProGlyProGlyProAlaGlyGluArgGlyPro	201
Db	1091	GGAGAACGGGCGAGAGCTGGGAGCTGGTGAGGCGGCCCTCTGGAGGCCAGGCGTC	1150
Qy	202	IleGlyProAlaGlyProProGlyGluArgGlyLysGlySerLysGlySer-----	219
Db	1151	CCTGGAGATGCTGGCATGCTGGGAGGCGGTGAGGCTGGCCACCGGGCTCAGCGGG	1210
Qy	220	-----GlnGlyProLys	223
Db	1211	GCCTCGGGCCACAGGGCCCTCCGGAGCCCTGGTGTCCGAGGCTCCAGGGCCAGAAG	1270
Qy	224	GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro	243
Db	1271	GGCAGCATGGAGACCCCGGCCCTCCAGGCCCCAGGCCCTCCGAGGTGAGCTGGGCGAC	1330
Qy	244	ProGlyProProGlyLysGlyLeuProGlyProGlnGlyProProGlyPheGlnGly	263
Db	1331	CGGGGTCCGGGAGTGCCGAAGGCCCTAAGGGAGACCAAGGTATTGACGGTTCCGACGGT	1390
Qy	264	LeuGlnGlyThrValGlyGlu-----	270
Db	1391	CTTCTGGGGATAAAGGAGAACTGGGTCCCAAGCGGCTGTGCGACCCCAAGGAGAGTCT	1450
Qy	270	-----	270
Db	1451	GGCAGTCGAGGGAGCTGGGCCCCAAAGCACCCAGGGTCCCAAGGCCACCACCGCGTGT	1510
Qy	271	-----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly	287
Db	1511	CAGGGTGTCGCCGGGCCCGCGTCTCTGGGCTGTGAGGCGCTCCCGGGTGTTCTCTGC	1570
Qy	288	Met-----	288
Db	1571	ATCACGGGGAAGCCGGAGTTCCGGGGAGGAGGCCACGACGACGCATCATGGAGCTG	1630
Qy	288	-----	288
Db	1631	TGTTGGGGGATGATCAGCGAACAAATTGCACAGTTAGCCGGCCACCTAAGGAAGCCTTTG	1690
Qy	289	-----ProGlyProLysGlyProProGlyProProGlyProSer	301
Db	1691	GCACCCGGGTCCATGTTGGTCCCGCTCAGCTGSCCCCTGGGCCCCAGGACCCCA	1750
Qy	302	GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro-----	313
Db	1751	GGCTCATGGTACACCTGGCTCGAGACCCCCCGGATACCCGCGTCCCACTGGGGAG	1810
Qy	314	-----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs	331
Db	1811	CTGGGAGACCCGGGCCACAG-----	1832
Qy	331	pLysCysTyrtyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG1	351
Db	1833	-----GAAACCAAGGTGCACAGCAGACAAG-----	1859

Alignment Scores:		
Pred. No.:	7, 77e-16	432
Score:	412.50	80
Percent Similarity:	55.70%	Conservative: 3
Best Local Similarity:	53.69%	Mismatches: 57
Query Match:	16.37%	Indels: 9
DB:	1	Gaps: 3

Alignment Scores:

```
US-09-763-712A-2_COPY_91_547 (1-457) x US-08-642-255-48 (1-432)
QY 158 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 177
Db 10 GCCCCAGCAGGTCGCGAGGCGCGCATGCCCCAGCAGGCGCCGAAAGGTCCGCTGGACCG 69
QY 178 ThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProProGlyProGlyProAlaGly 197
Db 70 GCTGTTCCACCGGGTGCTCCGGAGCTGCAGGCGCGCCAGGTGCGCTGGACCGGTGT 129
QY 198 GluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLys 217
Db 130 CCACCGGGTGCTCCGGAGCTGCAGGCGCGCCAGGTGCGCTGGACCGGTGTCCACCG 189
QY 218 GlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 237
Db 190 GGTGCTCCGGACCT-----GCAGGCGCGCCAGGTGCGCTGGACCGGTGTCCCA 240
QY 238 SerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGly 257
Db 241 CCGGTGCTCCGGAGCTGCAGGCGCGCCCA-----GGTCCGCTGGACCGGTGT 291
QY 258 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 277
Db 292 CCACCG-----GGTGCTCCGGAGCTGCAGGCGCGCCAGGTGCGCTGGACCGGT 342
QY 278 GlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyPro 297
Db 343 GGTCCACCGGGTGCTCCGGAGCTGCAGGCGCGCCAGGTGCGCTGGACCGGTGTCC 402
QY 298 ProGlyProSerGlyAlaValPro 306
Db 403 AAAGAGCTCACGGTCCGCGAGGTCCG 429

RESULT 10
US-08-159-784-4
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4
Alignment Scores:
Pred. No.: 1,03e-14 Length: 3394
Score: 409.00 Matches: 86
Percent Similarity: 48.98% Conservative: 10
Best Local Similarity: 43.88% Mismatches: 42
Query Match: 16.23% Indels: 58
DB: 1 Gaps: 7

US-09-763-712A-2_COPY_91_547 (1-457) x US-08-159-784-4 (1-3394)
QY 156 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySer----- 171
Db 553 ATGCCCGCGCGCCCGCAGGACCTCCAGGCGCGCCCGCCCTCCAGGACCTCTCTTTAGGAC 612
QY 172 -----GlnGlyProProGlyProGlyProGlyProGlyProGlyProGlyPro 182
Db 613 AGCAATGTGTTTGTCTGAGTCCAGCGCGCGCGCGCGCTCCAGGATTGCCAGGGAATCAG--- 669
QY 183 GlnLysGlyGluLysGlyGluProGlyProGlyProGlyProGlyProGlyProGlyPro 202
Db 670 -----GGCCCTCCAGCACCCCAAGCGGCCCAAGAGAGAGAGTG 705
QY 203 GlyProAlaGlyProProGlyGlu----- 210
Db 706 GSCCGCGCGCGGACCAACCGGCGAGTTTCGTTGACTTTTTCAGAAAGGAGGCTGAAATG 765
QY 211 ArgGlyGlyLysGlySerLysGlySerGlnGlyProGlyProLysGlySerArgGlySer----- 228
Db 766 AAGGGGAGAGAGGAGACCGAGGTGATGACGAGACAGAAAGGCGAAGGGGGAGGCGCGG 825
QY 229 -----ProGlyLysProGlyProGlyProGlyProGlyProGlyProGlyPro 239
Db 826 GCGCGCGGTTCCTCGGCTCCAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 885
QY 240 AspProGlyProProGlyProGlyLysGlu-----GlyLeuProGly 254
Db 886 TACCCTGGGATTCAGGTCCCAAGGAGAGAGACATCGGGGCGCAGCGCGCGCGCGCGCG 945
QY 255 ProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 274
Db 946 CCTCAGGACCGCGCGCATC---GGCTACGAGGGCGCGCGCGCGCGCGCGCGCGCGCG 1002
QY 275 GlyProArgGlyLeuProGlyLeuProGly-----ValProGly 287
Db 1003 GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
QY 288 MetProGlyProLysGlyProProGlyProProGlyProProGlyProSerGlyAla 303
Db 1063 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110

RESULT 11
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
```

```

; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621,091G
; FILING DATE: 11/30/90
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA: No. 5424408 applicable
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard S.
; REGISTRATION NUMBER: 28180
; REFERENCE/DOCKET NUMBER: 900983/RB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 972-1400
; TELEFAX: (212) 370-1622
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; STRAIN: Unknown
; INDIVIDUAL ISOLATE: Unknown
; DEVELOPMENTAL STAGE: Unknown
; CELL TYPE: Whole kidney
; IMMEDIATE SOURCE:
; LIBRARY: Bovine lens cDNA
; CLONE: KMC15
; POSITION IN GENOME: No. 5424408 known
; CHROMOSOME/SEGMENT: No. 5424408 known
; PUBLICATION INFORMATION: No. 5424408e
; US-07-621-091G-1

```

```

Alignment Scores:
Pred. No.: 9,63e-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
DB: 1 Gaps: 8

```

US-09-763-712a-2_COPY_91_547 (1-457) x US-07-621-091G-1 (1-1416)

```

QY 157 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAACCCAGGCATCGCTGCTCTCTGA 171

QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGluGlyGluProGly 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 CCTCAGGGAGTCTGTAGAGGCATTAAGGAGACAAGGGGTGTATGGAGAGCTGCG 231

QY 192 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 CAAAGAGTCCACCTGGAGCTATAGACATGGGGTCCACAGGTCATCCGGGAGCACA 291

QY 212 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 GGTGTCCCGGTCACCCAGGGCCAGAGGTGATCTGATGATTTCCAGGCATG 351

QY 232 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 AAAGGAGAGAGGGTAATTTCAGGATTTCCAGGACACCTGGACCTCCAGGGCAAAAGTGA 411

QY 252 -----LeuPro 253
   |||||
Db 412 CAAAAGGAGACCTGGAGTAGCTGGAGAGCCTGGCAGATGAAGATCATCTCCCTTCCA 471

```

```

QY 254 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 GGAAGCCAGGCCACCTGGTTTCAGCTGGAGAACACAGGGATGCAAGGAGAACCCGGGCC 531

QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 CCAGGACCCAGGAGATCCAGGACCTGTGGGCCAAAGGTAACACCGGGAGGATGTT 591

QY 294 ProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 592 CCACGAGGAACTCTGGACCACTGGAGAAAAGCAACAAGGTTGTAAGGAGAGACA 651

QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 652 GGACCA--CCTGGATCCGATCCGTCGCGCCAGGCTTGAAGGGGAAAC--TGGAGACACTGG 707

QY 330 ThrAspLysCysTyrrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 ACC---ACCTGCAGCAGGGGCGATGATGAGGGGCTTGTCTTTACCCGCGACAG----- 758

QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 369
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 759 ---CCAGACCCAGCAATTCCTCTGTCAGAGGGGACAGCCGCTCTATAGTGGGT 815

QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 816 TTCTCTCTCTTGTGTACAGGAATGAACAAGCCCATGACACGGA----- 860

QY 388 ArgGluAsnGluTrpLys 393
   ||||| |||||
Db 861 CCTGGGAACACTTGGCAG 878

RESULT 12
US-08-399-889-1
; Sequence 1, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-08-399-889-1

Alignment Scores:
Pred. No.: 9,63e-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
DB: 2 Gaps: 8

US-09-763-712a-2_COPY_91_547 (1-457) x US-08-399-889-1 (1-1416)

QY 157 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAACCCAGGCATCGCTGCTCTCTGA 171

QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGluGlyGluProGly 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 CCTCAGGGAGTCTGTAGAGGCATTAAGGAGACAAGGGGTGTATGGAGAGCTGCG 231

QY 192 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 211
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 CAAAGAGTCCACCTGGAGCTATAGACATGGGGTCCACAGGTCATCCGGGAGCACA 291

QY 212 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 GGTGTCCCGGTCACCCAGGGCCAGAGGTGATCTGATGATTTCCAGGCATG 351

QY 232 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 AAAGGAGAGAGGGTAATTTCAGGATTTCCAGGACACCTGGACCTCCAGGGCAAAAGTGA 411

QY 252 -----LeuPro 253
   |||||
Db 412 CAAAAGGAGACCTGGAGTAGCTGGAGAGCCTGGCAGATGAAGATCATCTCCCTTCCA 471

```

```
Db 172 CCTCCAGGAGTCTCTAGAACGCATATAAAGGAGACAAAGGGGTGTGATGGAGAGCCTGGC 231
QY 192 ProProGlyProAlaGlyLeuArgGlyProIleGlyProAlaGlyProGlyGluArg 211
Db 232 CAAAGAGGTCCACCTGGAGCTATAGGAGATAGGGGTACACAGGTATCCGGGAGCACCA 291
QY 212 GlyGlyLysGlySerGlyGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTCCAGCAGGAGCCAGAGGTATCTCTGGATTTCTATGATTTCAGGATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspGlyProGlyProGlyProGlyLysGluGly 251
Db 352 AAAGGGAAGAGGTAAATTCAGATTTCAGGACCACTGGACCACTCCAGGCAAGAGTGA 411
QY 252 -----LeuPro 253
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 254 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
Db 472 GGAAGCCCAAGGCCCACTGGTTCAGCTGGAGAACCCAGGATGCAAGGAGAACCCGGGCC 531
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 532 CCAGGACCACAGGAGATCCAGGACCTGTGGCCAAAAGGTAAACAGGGAGGATGGT 591
QY 294 ProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
Db 592 CCACCAAGAACTCTGGACCACTGGAGAAAAAGGCAACAAAGTTGTAAAGGAGAGCAA 651
QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
Db 652 GGACCA---CCTGGATCCAGTCCGCTGGAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 330 ThrAspLysCysTyrTyrPheSerValGluLysGluLysGluLysPheLeuPhe 349
Db 708 ACC---ACCTGACAGCGGCGAGTGTAGAGGGCTTGTCTTTACCCGGCACAG-----758
QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 369
Db 759 ---CCAGACCACAGCAATTCCTCTGTCCAGAGGACAGAGCCGCTCTATAGTGGT 815
QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
Db 816 TTCTCTTCTTTGTACAGGAAATGAACAAGCCCATGGACAGGA-----860
QY 388 ArgGluAsnGluTrpLys 393
Db 861 CCTGGGAACACTGGCAG 878

RESULT 13
US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
```

```
US-09-167-364-1
Alignment Scores:
Pred. No.: 9,63e-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
DB: 3 Gaps: 8

US-09-763-712a-2_COPY_91_547 (1-457) x US-09-167-364-1 (1-1416)
QY 157 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 176
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAAACAGGATGCCTGCTCTCTGGA 171
QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 191
Db 172 CCTCCAGGAGTCCCTGTAGAACGCATATAAAGGAGACAAAGGGTGTGATGGAGAGCCTGGC 231
QY 192 ProProGlyProAlaGlyLeuArgGlyProIleGlyProAlaGlyProGlyGluArg 211
Db 232 CAAAGAGGTCCACCTGGAGCTATAGGAGATAGGGGTACACAGGTATCCGGGAGCACCA 291
QY 212 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTCCAGCAGGAGCCAGAGGTATCTCTGGATTTCTATGATTTCAGGATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspGlyProGlyProGlyProGlyLysGluGly 251
Db 352 AAAGGGAAGAGGTAAATTCAGATTTCAGGACCACTGGACCACTCCAGGCAAGAGTGA 411
QY 252 -----LeuPro 253
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 254 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
Db 472 GGAAGCCCAAGGCCCACTGGTTCAGCTGGAGAACCCAGGATGCAAGGAGAACCCGGGCC 531
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 532 CCAGGACCACAGGAGATCCAGGACCTGTGGCCAAAAGGTAAACAGGGAGGATGGT 591
QY 294 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
Db 592 CCACCAAGAACTCTGGACCACTGGAGAAAAAGGCAACAAAGTTGTAAAGGAGAGCAA 651
QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
Db 652 GGACCA---CCTGGATCCAGTCCGCTGGAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 330 ThrAspLysCysTyrTyrPheSerValGluLysGluLysGluLysPheLeuPhe 349
Db 708 ACC---ACCTGACAGCGGCGAGTGTAGAGGGCTTGTCTTTACCCGGCACAG-----758
QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 369
Db 759 ---CCAGACCACAGCAATTCCTCTGTCCAGAGGACAGAGCCGCTCTATAGTGGT 815
QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
Db 816 TTCTCTTCTTTGTACAGGAAATGAACAAGCCCATGGACAGGA-----860
QY 388 ArgGluAsnGluTrpLys 393
Db 861 CCTGGGAACACTGGCAG 878

RESULT 14
US-09-439-897-1
; Sequence 1, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
```


QY 302 GlyAlaValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaProGluAsp 319
Db 919 GGCCAAAGAGGGCCCTTGGGGCTGCCAGGAATCCAGGGCGTGAAGGTCTGCCTGGTGAT 978
QY 320 AsnGlyCysPro 323
Db 979 AGAGGGGACCCCT 990

Search completed: March 21, 2003, 12:34:17
Job time : 61.552 secs

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Alignment Scores:

Pred. No.: 4,13e-130 Length: 2930
Score: 2514.00 Matches: 456
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x US-09-745-763-198 (1-2930)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 921 ATGAACAGCAGCTCAACATCATTCACAGGTGAGAGAACATCACCCTCTCTCAA 980
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnThr 40
Db 981 GCCACGAGCAGAACTGAAGACCTGCAGGACTTACAAAGATGCAGAGAAATGAACA 1040
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 1041 GCCATCAAGTTCAACCACTGAGGAGAACGCTCCAGCTCTTTGAGACGATATTGGAAC 1100
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
Db 1101 ATCATAGCAATATCAGTTACACAGCCACCACTCGGACGCTGACCAGCAATCTAAT 1160
QY 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 1161 GAAGTCAGGACCACTTGCACATACCTTACAAACACACAGATGATCTGACCTCTTG 1220
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 1221 AATAATACCTGGCCAAACATCGTTGGATTCTCTCTCAGGATGCAACAAGATTG 1280
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
Db 1281 ATGAGTCGAGGTGACACTGAAGTAGCACTTATCATCTAGTATTATGAAGAAATGAAG 1340
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
Db 1341 CTAGTAGACTCCAGCATGGTCAGTCTCATCAAGATTTTACAACTACTACAGGTCCACG 1400
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
Db 1401 GGCCCCAGGGTCCAAGAGGTGACAGAGGATCCAGGGGACCCCTGGCCCAACTGGCAAC 1460
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
Db 1461 AAGGGACAGAAAGAGAGAGGGGAGCGCTGGACCACCTTGGCCCTCGGGGTGAGAGAGCG 1520
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
Db 1521 CCATTTGGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAGGCTCCACG 1580

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
Db 1581 GGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAGGCCCCAGTGGGAC 1640
QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProGly 260
Db 1641 CCAGGCCCCCGGGCCACACAGGCAAGAGGACTCCCGGGCCCTCAGGGCCCTCCTGCG 1700
QY 261 PheGlnGlyLeuGlnGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1701 TTCCAGGAGCTTCAGGGCACCTTGGGAGCGCTGGGGTGCCTGGACCTCGGGACTGCCA 1760
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
Db 1761 GGCTTGCCTGGGTACACAGGATGCCAGGCCCAAGGGCCCCCGGCCCTCCTGGCCCA 1820
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1821 TCAGGAGCGGTGCTGCCCTGGCCCTGCAGAAATGAGCCAACCCCGCACGAGGACAAT 1880
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrPheSerValGluLys 340
Db 1881 AGCTGCCCGCCCTCCTGGGAAGAACTTCACAGACAATGCTACTATTTTCAGTTGAGAAA 1940
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
Db 1941 GAAATTTTGGAGATGCAAGCTTTCTGTGAAGACAAGCTTTCACATCTCTTTTCATA 2000
QY 361 AsnThrArgGluGlnGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
Db 2001 AACACTAGAGAGAACAGCAATGGATAAATAAACAAGATGGTAGGAGAGAGAGCCACTGG 2060
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
Db 2061 ATCGGCTTCACAGACTCAGAGCGTGAAATGAATGAAGTGGCTGGATGGGACATCTCCA 2120
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
Db 2121 GACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGCCAGCA 2180
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 2181 GAAGACTGTGCTGGGTGATTATCTGGGCAGTGGAGCATTTTCCAATGTGAAGACGTC 2240
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
Db 2241 AATACTTCATTTGGAAAAAGACAGGAGACAGTACTGTCTCATCTGCATTA 2291

RESULT 2

US-09-954-456-2274
; Sequence 2274, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711

[illegible]

Db	4632	CTGTGGTCTCTGCTGGTCCCTTAGCTGCACCTGGTCTCTCCAGGCTTACCAGGCTCTCAAGGC	4632
Qy	292	-----	292
Db	4692	CCAAAGGTAACAAAGCTCTACTGCAGCCGCTGCCACAAAGGTCACAGTGGCTTTCCA	4751
Qy	293	GlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGlu	312
Db	4752	GGGCTCTGGGCTCCAGGTCCACCTGGTGAAGTCATT-----CAG	4793
Qy	313	ProThrPro-AlaProGluAspAsnGlyCysProProHIStrPlyAsnPhetrAspLy	332
Db	4794	CCATTACCAATCF-----TGTCCTCCAAAAACGA-----	4824
Qy	332	sCysTyTyTyPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAs	352
Db	4825	-----GAAGACATACCTCAAGGCATGCAAG--	4848
Qy	352	pLysSerSerHisLeuValPheIleAsnThrArgGluGluGln-----	362
Db	4849	-CAGATCGACATGATAATATTCTTGATTACTCGGATGGAATGGAAGAAATATTTGGTTCC	4907
Qy	367	-----GlnTrpIleLysLysGlnMe	373
Db	4908	CTCAATTCCTGAAACAAGACATCGAGCATATGAAATTTCAATGGGTACTCAGACCAAT	4967
Qy	373	tValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLy	393
Db	4968	CCA-GCCCGAAGCTGTAAGACAGCTGCAACTGACCATCGCCATCTGACTTCCAGATGGTGGAATA	5026
Qy	393	strpLeuAsp-----GlyThrSerProAsp-----TyLysAsnTrpLy	406
Db	5027	TTGGATTGATCCTAACCAAGGTGGCTCAGGAGATCCTTTCAAAGTTTACTGTAATTTCCAC	5086
Qy	406	sAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLe	426
Db	5087	ATCTGGT-----GGTGAGACTTGC-----	5105
Qy	426	uileTyAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhelIleCysG	446
Db	5106	-ATTATCCA-----GACAAAAAATCTGAGGGAGTGAAG-AATTTCAATCATGGC	5151
Qy	446	uLys 447	
Db	5152	CAA 5155	
RESULT 6			
US-09-954-456-762			
; Sequence 762, Application US/09954456			
; Patent No. US20020115057A1			
; GENERAL INFORMATION:			
; APPLICANT: Young, Paul			
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agen			
; FILE REFERENCE: 689290-76			
; CURRENT APPLICATION NUMBER: US/09/954,456			
; CURRENT FILING DATE: 2001-09-18			
; PRIORITY APPLICATION NUMBER: US/60/233,617			
; PRIORITY FILING DATE: 2000-09-18			
; PRIORITY APPLICATION NUMBER: US/60/234,052			
; PRIORITY FILING DATE: 2000-09-20			
; PRIORITY APPLICATION NUMBER: US/60/234,923			
; PRIORITY FILING DATE: 2000-09-25			
; PRIORITY APPLICATION NUMBER: US/60/235,134			
; PRIORITY FILING DATE: 2000-09-25			
; PRIORITY APPLICATION NUMBER: US/60/235,637			
; PRIORITY FILING DATE: 2000-09-26			
; PRIORITY APPLICATION NUMBER: US/60/235,638			
; PRIORITY FILING DATE: 2000-09-26			
; PRIORITY APPLICATION NUMBER: US/60/235,711			
; PRIORITY FILING DATE: 2000-09-27			
; PRIORITY APPLICATION NUMBER: US/60/235,720			
; PRIORITY FILING DATE: 2000-09-27			

Qy	293	GlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGlu	312
Db	4752	GGGCGCTCCTGGGCTCCAGGTCCACCTGGTGAAGTCATT	4793
Qy	313	ProThrPro-AlaProGluAspAsnGlyCysProHisHisTrpLysAsnPhThrAspLys	332
Db	4794	CGTTTACCAATCT	4824
Qy	332	sCysTyTyPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAs	352
Db	4825		4848
Qy	352	plySerSerHisLeuValPheIleAsnThrArgGluGluGln	366
Db	4849	-CAGATGCAGATGATAATATCTTGATTACTCGGATGGAATGGAAGAAATATTTGGTTCC	4907
Qy	367		373
Db	4908	CTCAATTCCCTGAACAAGACATCGAGCATATGAAATTTCCAATGGTACTCAGACCAAT	4967
Qy	373	tValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys	393
Db	4968	CCA-GCCCGAACTGTAAAGACCTGCAACTCAGCCATCTCTGACTTCCAGATGGTGAATA	5026
Qy	393	sTrpLeuAsp	406
Db	5027	TTGGATTGATCCTTAACCAAGTTTGCTCAGGAGATTCTTCAAAGTTTACTGTAAATTTCCAC	5086
Qy	406	sAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLe	426
Db	5087	ATCTGGT	5105
Qy	426	uIleTyAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhIleCysG	446
Db	5106	-ATTTATCCA	5151
Qy	446	uLys	447
Db	5152	CAMA	5155
RESULT 7			
US-09-954-456-782			
; Sequence 782, Application US/09954456			
; Patent No. US20020115057A1			
; GENERAL INFORMATION:			
; APPLICANT: Young, Paul			
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agen			
; TITLE OF INVENTION: Sets			
; FILE REFERENCE: 689290-76			
; CURRENT APPLICATION NUMBER: US/09/954,456			
; CURRENT FILING DATE: 2001-09-18			
; PRIORITY APPLICATION NUMBER: US/60/233,617			
; PRIORITY FILING DATE: 2000-09-18			
; PRIORITY APPLICATION NUMBER: US/60/234,052			
; PRIORITY FILING DATE: 2000-09-20			
; PRIORITY APPLICATION NUMBER: US/60/234,923			
; PRIORITY FILING DATE: 2000-09-25			
; PRIORITY APPLICATION NUMBER: US/60/235,134			
; PRIORITY FILING DATE: 2000-09-25			
; PRIORITY APPLICATION NUMBER: US/60/235,637			
; PRIORITY FILING DATE: 2000-09-26			
; PRIORITY APPLICATION NUMBER: US/60/235,638			
; PRIORITY FILING DATE: 2000-09-26			
; PRIORITY APPLICATION NUMBER: US/60/235,711			
; PRIORITY FILING DATE: 2000-09-27			
; PRIORITY APPLICATION NUMBER: US/60/235,720			
; PRIORITY FILING DATE: 2000-09-27			
; PRIORITY APPLICATION NUMBER: US/60/235,840			
; PRIORITY FILING DATE: 2000-09-27			
; PRIORITY APPLICATION NUMBER: US/60/235,863			
; PRIORITY FILING DATE: 2000-09-27			
; NUMBER OF SEQ ID NOS: 2276			
; SOFTWARE: PatentIn version 3.0			

```

; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-78

```

Alignment Scores:

Pred. No.:	1,11e-16	Length:	6728
Score:	452.50	Matches:	99
Percent Similarity:	52.11%	Conservative:	12
Best Local Similarity:	46.48%	Mismatches:	53
Query Match:	17.96%	Indels:	49
DB:	10	Gaps:	8

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-954-456-782 (1-6728)

QY	156	LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	175
Db	1485	GTTCAGGACCCCTGGCCCTCTCGAGGAGAAGAACGAGGAGCTCGAGGTGAACCC	1544
QY	176	GlyProThrGlyAsnLysGlnLysGlyGluLysGlyGluPro-	190
Db	1545	GGACCCACTGGCTGCTCCCGGACCCCTGGCAGGCGTGTGACCTGGTAGCCGTGGTTTC	1604
QY	191	-----GlyProProGlyProAlaGlyGluArgGlyProIleGly	203
Db	1605	CTTGGCGCAGATGGTGTCTGCTGCCAACAGGCTCCGCTGGTGAACGTGGTTCTCTCGTGC	1664
QY	204	ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln	220
Db	1665	CCGCGTGGCCCCAAAGAGTCTCTCGTGAAGCTGGCTGCCGCGAAGCTGGTCTCGCT	1724
QY	221	GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp	240
Db	1725	GGTGCCAAAGGCTGACTGGAAGCCCTGGCAGCCCTGCTCTGATGCGCAAACT-	1778
QY	241	ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProPro-	259
Db	1779	---GGCCCCCTGGTCCCGCGGTCAAGATGTCGCCCGGACCCCGAGGCCACCTGGT	1835
QY	260	-----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro	271
Db	1836	GCCCCGTGTACGGCTGGTGTATGGATTCCTGGACCTAAAGGTGCTGTGGAGACCC	1895
QY	272	GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal-	285
Db	1896	GGCAAGGCTGGAGGAGGAGTGTTCGCGACCCCTGGCGCTGCTGGTGGCAAA	1955
QY	286	ProGlyMetProGlyProLysGlyProProGlyProProGlyProSerGly-	302
Db	1956	GATGGAGGCGTGGAGCTCAGGAGCCCTGGCCCTGCTGGTCTGCCGCTGGCGGAGAGGT	2015
QY	303	-----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla-	316
Db	2016	GAACAAGGCCCTGCTGGCTCCCCC---GGATTCCAGGCTCTCCCTGGTCTGCTGGTCCCT	2072
QY	317	-----ProGluAspAsnGlyCysPro	323
Db	2073	CCAGGTGAAGCAGGCAAACTCGTGAACAGGCGTGTCT	2111

RESULT 8

```

US-09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14

```

```

: PRIOR APPLICATION NUMBER: US 60/211,379
:
: PRIOR FILING DATE: 2000-06-14
: PRIOR APPLICATION NUMBER: US 60/237,054
: PRIOR FILING DATE: 2000-10-02
: NUMBER OF SEQ ID NOS: 3950
: SOFTWARE: Patent In ver. 2.1

```

```

; SEQ ID NO 3946
: LENGTH: 6728

```

```

; LENGTH: 6726
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274615
US-09-880-107-3946

```

Alignment Scores:

Argument Scores:					
Pred. No.:	1,11e-16	Length:	6728		
Score:	452.50	Matches:	99		
Percent Similarity:	52.11%	Conservative:	12		
Best Local Similarity:	46.48%	Mismatches:	53		
Query Match:	17.96%	Indels:	49		
DB:	10	Gaps:	8		

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-880-107-3946 (1-6728)

156	LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro	175
QY	:::	
1485	GTTCAAGGACCCCTCTGGAGAGGAAGAAAGAGGAGCTCGAGGTGAACCC	1544
Db		
176	GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro-	190
QY		
1545	GGACCACTGGCCTGCCCGGACCCCTGCGGAGCGTGGTGACCTGGTAGCCGTGGTTC	1604
Db		
191	-----GlyProProGlyProAlaGlyGluArgGlyProIleGly	203
QY		
1605	CCTGGCGCAGATGTTGCTGCTCCCAAGGTCCCGCTGGTGAACGTGTTCTCTGGC	1664
Db		
204	ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln	220
QY		
1665	CCCGCTGGCCCCAAGAGTCTCTGTGTGAAGCTGGTGTCCGGGTGAAGCTGGTCTGCCT	1724
Db		
221	GlyProLysGlySerArgGlySerProGlyLysProGlyLysProGlnGlyProSerGlyAsp	240
QY		
1725	GGTGCAAGGCTGTGACTGGAAGCCCTGCGACCCCTGGTCTGTATGGCAAACT-	1778
Db		
241	ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProPro-	259
QY		
1779	---GGCCCCCTGTGTCGCCGTCAAGATGTGTGCCCGCGGACCCCAAGGCCACTGGT	1835
Db		
260	-----GlyPheGlnGlyLeuGlnGlyThrValGlyGlyClnPro	271
QY		
1836	GCCCCGTGTCAGCTGGTGGTATGGATTCCTCGAACCTAAAGTGTGCTGGAGAGCC	1895
Db		
272	GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal-	285
QY		
1896	GGCAAGCTGGAGAGGAGGTGTTCCGAGCCCTCCGCGCTGTCGGTCTGCTGGCAAA	1955
Db		
286	ProGlyMetProGlyProLysGlyProProGlyProProGlyProSerGly-	302
QY		
1956	GATGGAGAGGCTGGAGCTCAGGAGCCCCCTGCGCCTGTGTGGTCCGCTGGCGAGAGAGT	2015
Db		
303	-----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla	316
QY		
2016	GAACAAGCCCTGCTGGCTCCCC---GGATTCCAGGGTCTCCCTGGTCTGCTGGTCT	2072
Db		
317	-----ProGluAspAsnGlyCysPro	323
QY		
2073	CCAGGTGAAGCAGGCAAACTGCTGAACAGGTTGTCCT	2111
Db		

RESULT 9

US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:

```
: APPLICANT: Young, Paul
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cano
: FILE REFERENCE: 689290-76
: CURRENT APPLICATION NUMBER: US/09/954,456
: CURRENT FILING DATE: 2001-09-18
: PRIOR APPLICATION NUMBER: US/60/233,617
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234,052
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,923
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/235,134
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/235,637
: PRIOR FILING DATE: 2000-09-26
: PRIOR APPLICATION NUMBER: US/60/235,638
: PRIOR FILING DATE: 2000-09-26
: PRIOR APPLICATION NUMBER: US/60/235,711
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,720
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,840
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,863
: PRIOR FILING DATE: 2000-09-27
: NUMBER OF SEQ ID NOS: 2276
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 786
: LENGTH: 5416
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-954-456-786

Alignment Scores:
Pred. No.:          2,67e-16      Length:         5416
Score:              444.00        Matches:         115
Percent Similarity: 43.40%       Conservative:    23
Best Local Similarity: 36.16%     Mismatches:     79
Query Match:        17.62%       Indels:         102
DB:                 10           Gaps:           10

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-954-456-786 (1-5416)

Qy   107   IleArgLeuAsnSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAsp   126
::: |||:::|||||||
Db   2280 GTCTATTGGAGCGGAGGTCTTCTTGACCCCGCCTGATGGAAACAAGGGTGAAC   2339

Qy   127   ThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer----- 144
||| :::::: :::: |
Db   2340 CTGGTGTGGTTGGTCGTGGGCACACTGCTGGTCCATCTGCTAGTGAGACTCCCAGGAG   2399

Qy   145 -----LysHisGlyGlnLeuLeuLysAsn   152
||| |||:::
Db   2400 AGAGGGTGTCTGTCGCATACCTGGAGGCCAAGGAGGAGTGAACCTGGTCTCAGAG   2459

Qy   153 -----PheThrIleLeuGlnGlyPro   159
::: |||
Db   2460 GTGAATAATTGTAACCCTGGCAGAGATGGTCTGTGGTGTCTCATGGTGTCTGTA--GGTGC   2518

Qy   160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProProGly   176
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   2519 CCTGGTCTCTGCGAGCACAGGTGACCGGGGCGAAGCTGGGCTGCTGGTCTGCTGTGGT   2578

Qy   177 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla   196
||| ||| ::||| |||||:||||| ||||| ||||| |||||
Db   2579 CCGTCTGCTGTCGGGAAGCCCTGCTGTGAACGTGGCAGGCTGCTCTGCTGGCCCCAAC   2638

Qy   197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly   213
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   2639 GGATTTCTGTGCGGCTGTGCTGCTGGTCAACCGGGGTCTTAAGGAGAAGAGGAGGC   2698

Qy   214 LysGlySerLysGlySerGln-----                220
```

Db	2699	AAAGGGCCTAAGGGTGAACCGGTGTTGGTCCCAAGGCCGGTTTGAGACTGCTGGC	2758
Qy	221	-----GlyProLysGlySerArgGlySer- 	228
Db	2759	CCAGCTGGTCCAATGTGTCCTCCCCCGGTCTGCTGGAAGTGTGGTGGATGGAGGCCCCCCT	2818
Qy	229	-----ProGlyLysProGlyProGlnGlyPro 	237
Db	2819	GGTATGACTGGTTTCCTGGTCTGGAGGAGACTGGTCCCCCAGGACCCTCTGGTATT	2878
Qy	238	SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln--- 	256
Db	2879	TCTGGCCCTCTGCTGCTCCCTGGTCTGCTGGGAAAGAAGGGCTTCGTGGTCTCGTGGT	2938
Qy	257	-----GlyProProGlyPheGln 	262
Db	2939	GACCAAGGTCCAGTTGGCCGAAC TGGAAGTAGGTGCAGTTGGTCCCCCTGGCTCGCT	2998
Qy	263	GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro :	276
Db	2999	GGTGAGAGGGTCCCTCTGGAGAGGCTGGTACTGTGGACCTCTGGCACCTCCAGGTCCT	3058
Qy	277	ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGly ::: :: :: :: ::	296
Db	3059	CAGGCTCTCTTGGTGCTCCTGTATTCGTGGCTCCTCGGTGCGAGAGGTGAACGTGGT	3118
Qy	297	ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro :: ::	313
Db	3119	CTACCTGGTGTGCTGGTGTGGGTGAACCTGGTCTCTTGGCATTTGCCGGCCCTCCT	3178
Qy	314	-----ThrProAlaProGluAspAsnGlyCysPro :::	323
Db	3179	GGGCGCGTGGTCTCTCTGGTGTGGTAGTCTCTGGAGTCAACGGTGTCTCT	3232
RESULT 10			
US-09-880-107-2094			
; Sequence 2094, Application US/09880107			
; Patent No. US20020142981A1			
; GENERAL INFORMATION:			
; APPLICANT: Horne, Darci T.			
; APPLICANT: Vockley, Joseph G.			
; APPLICANT: Scherf, Uwe			
; APPLICANT: Gene Logic, Inc.			
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer			
; FILE REFERENCE: 44921-5028-WO			
; CURRENT APPLICATION NUMBER: US/09/880,107			
; CURRENT FILING DATE: 2001-06-14			
; PRIOR APPLICATION NUMBER: US 60/211,379			
; PRIOR FILING DATE: 2000-06-14			
; PRIOR APPLICATION NUMBER: US 60/237,054			
; PRIOR FILING DATE: 2000-10-02			
; NUMBER OF SEQ ID NOS: 3950			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 2094			
; LENGTH: 5416			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03464			
US-09-880-107-2094			
Alignment Scores:			
Pred. No.:	2,67e-16	Length:	5416
Score:	444.00	Matches:	115
Percent Similarity:	43.40%	Conservative:	23
Best Local Similarity:	36.16%	Mismatches:	79
Query Match:	17.62%	Indels:	102
DB:	10	Gaps:	10
US-09-763-712A-2_COPY_91_547 (1-457) x US-09-880-107-2094 (1-5416)			
Ov	107	IleArgLeuAspSerValSerLeuArqMetGlnGlnAspLeuMetArgSerArgLeuAsp	126

Qy 107 IleArqLeuAspSerValSerLeuArqMetGlnGlnAspLeuMetArqSerArqLeuAsp 126


```
Db 2280 GTCCATTATGGAGCGAGGTCCTCTTGGACCCCGAGGCGCTGATGAAACAGGCGTGAAC 2339
Qy 127 ThrGluValAlaAsnLysSerValIleMetGluGluMetLysLeuValAspSer----- 144
Db 2340 CTGGTGTGGTGGTCTGTGGCACTGCTGTGGTCCATCTGGTCCATCTGGTACCTCCAGAG 2399
Qy 145 -----LysHisGlyGlnLeuIleLysAsn 152
Db 2400 AGAGGGGTGCTGCTGGCATACCTGGAGGCAAGGAGAGAAAGGTGAACCTGGTCTCAGAG 2459
Qy 153 -----PheThrIleLeuGlnGlyPro 159
Db 2460 GTGAATTTGGTAACCTGGCAGAGATGCTGCTGGTGTCTCATGCTGCTGA-GGTGCC 2518
Qy 160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProProGly 176
Db 2519 CCTGTCTCTGTCGGAGCCACAGGTACCGGGCGAAGCTGGGGCTGTGGTCTCTCTGCT 2578
Qy 177 ProThrGlyAsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAla 196
Db 2579 CCGTGTGTCTCCCGGGAAGCCCTGTGAACGTGCGAGGTGCTGCTGGCCCCCAAC 2638
Qy 197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 213
Db 2639 GGATTGTGTGCTCGGCTGTGCTGCTGCTCAACCGGCTGCTTAAAGGAGAAAGAGGAGGC 2698
Qy 214 LysGlySerLysGlySerGln----- 220
Db 2699 AAAGGCGCTAAGGTTGAACACGGTGTGTGTGTCACACAGGCCCGCTGGAGCTGCTGGC 2758
Qy 221 -----GlyProLysGlySerArgGlySer----- 228
Db 2759 CCAGCTGTGTCCAAATGTTCCCGCGTCTCTGTGGAAGTGTGGTGTGATGGAGGCCCGCT 2818
Qy 229 -----ProGlyLysProGlyProGlnGlyPro 237
Db 2819 GGTATGACTGTTTCCCTGTGCTGCTGACGGAGTGTGCTCCCGGAGGACCTCTGTGTATT 2878
Qy 238 SerGlyAspProGlyProProGlyProGlyLysGluGlyLysGlyLysGlyProGln--- 256
Db 2879 TCTGGCCCTCTGGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2938
Qy 257 -----GlyProProGlyPheGln 262
Db 2939 GACCAAGTCTCAGTTGGCGAAGTGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2998
Qy 263 GlyLeuGln-----GlyThrValGlyLysGlyValProGlyPro 276
Db 2999 GGTGAGAAGGTCTCCTCTGGAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3058
Qy 277 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGly 296
Db 3059 CAGGGTCTCTTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3118
Qy 297 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 313
Db 3119 CTACCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3178
Qy 314 -----ThrProAlaProGluAsnGlyCysPro 323
Db 3179 GGGGCGCGGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3232
```

RESULT 11

US-10-044-090-104

; Sequence 104, Application US/10044090

; Patent No. US20020137081A1

; GENERAL INFORMATION:

; APPLICANT: Olga Bandman

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

; FILE REFERENCE: PA-0028 US

; CURRENT APPLICATION NUMBER: US/10/044,090

; CURRENT FILING DATE: 2002-01-09

```
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 104
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.19
; NAME/KEY: unsure
; LOCATION: 1492-1495, 1663, 1669, 4012, 4055
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-104
```

Alignment Scores:

Pred. No.:	3,27e-16	Length:	4149
Score:	440.50	Matches:	95
Percent Similarity:	49.32%	Conservative:	13
Best Local Similarity:	43.38%	Mismatches:	58
Query Match:	17.48%	Indels:	53
DB:	12	Gaps:	6

US-09-763-712A-2_COPY_91_547 (1-457) x US-10-044-090-104 (1-4149)

```
Qy 158 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 177
Db 1307 GGTCTGCTGTGCTCGGGGAAGCCCTGCTGAACGTGTGTGAGTGTGCTGCTGCTGCTGCT 1366
Qy 178 Thr-----Gly 179
Db 1367 AATGATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426
Qy 180 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAlaGlyGluArg 199
Db 1427 GCCAAGGCGCTAAGGTGAAACGGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1486
Qy 200 GlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyGlySerLysGlySer 219
Db 1487 GGGCCNNNGTCCAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1546
Qy 220 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro----- 237
Db 1547 CTGTGTATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1606
Qy 238 ---SerGlyAspProGlyProGlyProGlyProGlyLysGlyGlyLeuProGlyProGln 256
Db 1607 ATTTCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1666
Qy 257 -----GlyProProGlyPhe 261
Db 1667 GGNAGCAAGGACAGCAGCGCCGACCTGGAGAAAGTAGGAGCACCGGCTCCCGCTGCTGCT 1726
Qy 262 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 275
Db 1727 GCTGTGAGAAAGGTCCTCTCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1786
Qy 276 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 295
Db 1787 CTTAGGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1846
Qy 296 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 312
Db 1847 GCTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906
Qy 313 Pro-----ThrProAlaProGluAsnGlyCysPro 323
Db 1907 CTGGGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1963
```

RESULT 12

US-09-880-107-3947

; Sequence 3947, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Uwe
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3947
 ; LENGTH: 5086
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(5086)
 ; OTHER INFORMATION: n = a or c or g or t
 ; US-09-880-107-3947
 Alignment Scores:
 Pred. No.: 3,92e-16 Length: 5086
 Score: 440.50 Matches: 114
 Percent Similarity: 43.08% Conservative: 23
 Best Local Similarity: 35.85% Mismatches: 80
 Query Match: 17.48% Indels: 102
 DB: 10 Gaps: 10
 US-09-763-712A-2_COPY_91_547 (1-457) x US-09-880-107-3947 (1-5086)
 QY 107 IleArgLeuAspSerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAsp 126
 Db 1950 GTCTATTGGAAGCGAGGTCTCTTGACCCCGAGGCTGATGGAAACAAAGGTTGAAC 2009
 QY 127 ThrGluValAlaAsnLeuSerValIleMetGluMetLysLeuValAspSer----- 144
 Db 2010 CTGGTGGTGGTGGTGGGCACTGCTGCTCCTAGTCCATGTTGAGTGGACTCCAGGAG 2069
 QY 145 -----LysHisGlyGlnLeuIleLysAsn 152
 Db 2070 AGAGGGGTGCTGCTGCATACCTGGACGAAGGAGAAAGGTTGAACCTGCTCAGAG 2129
 QY 153 -----PheThrIleLeuGlnGlyPro 159
 Db 2130 GTGAATTTGTAACCTGGCGAGAGATGGTCTGCTGGTGTCTAGTGGTCTGTA-GGTGCC 2188
 QY 160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProProGly 176
 Db 2189 CCTGGTCTCTGGAGCCACAGGTGACGGGGCGAAGCTGGGGTCTGGTCTGCTGGT 2248
 QY 177 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAla 196
 Db 2249 CCTGCTGGTCTCTGGGGAAGCCCTGGTGAAGCTGGCGAGGTCTGCTGCTGGCCCAAC 2308
 QY 197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 213
 Db 2309 GGATTTGCTGCTGGCTGGTGGTCTGCTGGTCAACCGGGTGTAAAGGAGAAAGAGGACC 2368
 QY 214 LysGlySerLysGlySerGln----- 220
 Db 2369 AAAGGGCTAAGGGTGAAGAGCGGTGTTGGTCCACAGGCCCGCTGGAGCTGCTGGC 2428
 QY 221 -----GlyProLysGlySerArgGlySer----- 228
 Db 2429 CCAGCTGGTCCAAATGGTCCCGCGCTCTGCTGGAGTCTGTTGATGAGGAGGCCCCCT 2488
 QY 229 -----ProGlyLysProGlyProGlnGlyPro 237
 Db 2488

```
Db 2248 GGTCTGCTGGTCTCTGGGGAAGCCCTGGTGAACGTGGTGAAGTGGTGGTCTGGCCCTCTGGCCCT 2307
Qy 178 Thr-----Gly 179
Db 2308 AATGGATTGCTGGTCTCTGGTCTGGTCAACCTGGTGTAAAGGAGAAAGAGGA 2367
Qy 180 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArg 199
Db 2368 GCCAAAGGCCCTAAAGGCTGAAACCGGTGTGTGGTCCACAGGCCCTGGTGGAGCTGCT 2427
Qy 200 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlyGlySerLysGlySer 219
Db 2428 GGCACAGCTGGTCCAAATGGTCCCGGCTGGTGTGAAGTGGTGGTGAAGGAGGCCCT 2487
Qy 220 GlnGlyProLysGlySerArgLysProGlyLysProGlyProGlyProGlyPro----- 237
Db 2488 CTTGGTATGACTGGTTTCCCTGGTGTGGACGACTGGTCCCGCCAGGACCTCTGGT 2547
Qy 238 ---SerGlyAspProGlyProGlyProGlyLysGlyGluGlyLysGlyProGlyProGln 256
Db 2548 ATTCTGGCCCTCTCTGGTCTGGTCCCGGCTGGTGTGAAGTGGTGGTGGTGGTGGTGGT 2607
Qy 257 -----GlyProGlyPhe 261
Db 2608 GTGACCAAGTCCAGTGGGCCAACTGGAGAAAGTAGGTGCAAGTGGTGGTGGTGGTGGTGGT 2667
Qy 262 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 275
Db 2668 GCTGGTGAGAGGGTCCCTCTGGAGAGGCTGGTACTGCTGGACCTCTGGGCACTCCAGGT 2727
Qy 276 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 295
Db 2728 CCTCAGAGGCTCTTGTGGTCTCTGGTATCTGGGTCTCCCTGGTCTCGAGAGGTGAACGT 2787
Qy 296 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 312
Db 2788 GGTCTACCAAGTGTGGTGGTGGTGAACCTGGTGAACCTGGTGGTGGTGGTGGTGGTGGT 2847
Qy 313 Pro-----ThrProAlaProGluAspAsnGlyCysPro 323
Db 2848 CTGGGGCCCTGGTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2904

RESULT 14
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
```

```
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.: 3,97e-16 Length: 5145
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 17.48% Indels: 53
DB: 10 Gaps: 6

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-925-299-206 (1-5145)
Qy 158 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyPro 177
Db 2248 GGTCTGCTGGTCTCTGGGGAAGCCCTGGTGAACGTGGTGAAGTGGTGGTCTGGCCCT 2307
Qy 178 Thr-----Gly 179
Db 2308 AATGGATTGCTGGTCTCTGGTGTCAACCTGGTGTAAAGGAGAAAGAGGA 2367
Qy 180 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArg 199
Db 2368 GCCAAAGGCCCTAAAGGCTGAAACCGGTGTGTGGTCCACAGGCCCTGGAGCTGCT 2427
Qy 200 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer 219
Db 2428 GGCACAGCTGGTCCAAATGGTCCCGGCTGGTGTGAAGTGGTGGTGGTGGTGGTGGTGGT 2487
Qy 220 GlnGlyProLysGlySerArgLysProGlyLysProGlyProGlyProGlyPro----- 237
Db 2488 CTTGGTATGACTGGTTTCCCTGGTGTGGACGACTGGTCCCGCCAGGACCTCTGGT 2547
Qy 238 ---SerGlyAspProGlyProGlyProGlyLysGlyGluGlyLysGlyProGlyProGln 256
Db 2548 ATTCTGGCCCTCTCTGGTCTGGTCCCGGCTGGTGTGAAGTGGTGGTGGTGGTGGTGGTGGT 2607
Qy 257 -----GlyProGlyPhe 261
Db 2608 GTGACCAAGTCCAGTGGGCCAACTGGAGAAAGTAGGTGCAAGTGGTGGTGGTGGTGGTGGT 2667
Qy 262 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 275
Db 2668 GCTGGTGAGAGGGTCCCTCTGGAGAGGCTGGTACTGCTGGACCTCTGGGCACTCCAGGT 2727
Qy 276 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 295
Db 2728 CCTCAGAGGCTCTTGTGGTCTCTGGTATCTGGGTCTCCCTGGTCTCGAGAGGTGAACGT 2787
Qy 296 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 312
Db 2788 GGTCTACCAAGTGTGGTGGTGGTGAACCTGGTGAACCTGGTGGTGGTGGTGGTGGTGGTGGT 2847
Qy 313 Pro-----ThrProAlaProGluAspAsnGlyCysPro 323
Db 2848 CTGGGGCCCTGGTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2904

RESULT 15
US-10-044-090-22
; Sequence 22, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 5432
; TYPE: DNA
```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
US-10-044-090-22

Alignment Scores:

Pred. No.:	4.16e-16	Length:	5432
Score:	440.50	Matches:	114
Percent Similarity:	43.08%	Conservative:	23
Best Local Similarity:	35.85%	Mismatches:	80
Query Match:	17.48%	Indels:	102
DB:	12	Gaps:	10

US-09-763-712A-2_COPY_91_547 (1-457) x US-10-044-090-22 (1-5432)

```
QY 107 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAsp 126
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2283 GTCTATTGGAGCGAGGTCCTTCGGACCCCGGCTGATGGAAACAAGGGTGAAC 2342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 ThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer----- 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2343 CTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 -----LysHisGlyGlnLeuLeuLysAsn 152
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2403 AGAGGGGTGCTGTGGCATACCTGGAGCAAGGAGAAAGGGTGAACCTGTCTCAGAG 2462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 -----PheThrIleLeuGlnGlyPro 159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2463 GTGAATTGGTAACTGGCAGAGATGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2521
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProGly 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2522 CCTGGTCTGTGGAGCCACAGGTACCGGGGAGAGCTGGGGCTGCTGTGGTCTGTGGTGTGGT 2581
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 177 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2582 CCTGTGTGGTCTCGGGGAAAGCCCTGGTGAACCTGGTGAACCTGGTGAACCTGGTGAACCT 2641
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2642 GGATTTGTGGTCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2701
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 214 LysGlySerLysGlySerGln-----ProGlyLysProGlyProGlnGlyPro 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2702 AAAGGCTTAAGGGTGAAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2761
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 221 -----GlyProLysGlySerArgGlySer----- 228
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2762 CCAGTGTGTCCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2821
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 229 -----ProGlyLysProGlyProGlnGlyPro 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2822 GGTATGACTGGTTTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2881
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln--- 256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2882 TCTGGCCCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 257 -----GlyProProGlyPheGln 262
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2942 GACCAAGTTCAGTTGGCGAAGTGGAGAGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 3001
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 263 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3002 GTTGAGAGGGTCCCTCTGTGGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3061
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 277 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3062 CAGGTCCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 297 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 313
```

```
Db 3122 CTACCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3181
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 314 -----ThrProAlaProGluAspAsnGlyCysPro 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3182 GGGGCCCCGTGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: March 21, 2003, 12:32:08
Job time : 248.049 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:08:41 ; Search time 1510.54 Seconds
(without alignments)
4899.781 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547
Perfect score: 2520
Sequence: 1 MNSQNSFTQGMENITTSQ.....EDVNFNICEKDRETVLSSAL 457

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlp
-O=/cgn2_1/USPROOPI/US09763712/runat_14032003_100950_18109/app_query.fasta_1.1877
-DB=EST -QFMT=tastap -SUFF1=1st -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn_1.1.4575 @runat_14032003_100950_18109 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-FGAPOP=10 -FGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2344	93.0	3305	11	BC009162	BC009162 Mus muscu
2	1460.5	58.0	957	14	BQ955927	BQ955927 AGENCOURT
3	1256	49.8	861	14	BQ713873	BQ713873 AGENCOURT
4	1251	49.6	906	13	B1456109	B1456109 603172765
5	1239.5	49.2	936	14	BQ891432	BQ891432 AGENCOURT
6	1022	40.6	580	12	BE910803	BE910803 601661855
7	1007	40.0	668	14	BQ674807	BQ674807 AGENCOURT
8	957	38.0	601	10	BE290299	BE290299 601089246
9	916.5	36.4	967	9	AL568743	AL568743 AL568743
10	842	33.4	552	10	AW958053	AW958053 EST370123
11	831	33.0	638	10	BB248064	BB248064 BB248064
12	798.5	31.7	884	14	BQ934501	BQ934501 AGENCOURT
13	767	30.4	808	14	BQ771366	BQ771366 UI-M-F10-
14	669	26.5	500	14	BM676508	BM676508 UI-E-EJ0-
15	663	26.3	507	14	BM713891	BM713891 UI-E-EJ0-
16	627.5	24.9	591	13	B1445884	B1445884 dai33d10.
17	601.5	23.9	1082	14	BM907108	BM907108 AGENCOURT
18	600	23.8	357	10	AW379436	AW379436 CM4-HT024
19	594	23.6	350	9	AA012704	AA012704 RPUI402CG
20	588	23.3	339	9	AA304251	AA304251 EST17006
21	560	22.2	517	9	AL543000	AL543000 AL543000
22	538	21.3	581	14	BQ127513	BQ127513 i160h05.Y
23	537	21.3	310	9	AA361740	AA361740 EST71069
24	516	20.5	465	10	AW240221	AW240221 up30b07.Y
25	503.5	20.0	839	13	BT697412	BT697412 603348154
26	478	19.0	943	14	BQ887163	BQ887163 AGENCOURT
27	477	18.9	679	14	W27610	W27610 35b6 Human
28	460	18.3	441	10	BB749743	BB749743 AB749743
29	458.5	18.2	569	10	AV616076	AV616076 AV616076
30	457.5	18.2	1029	17	CNS04VYG	CNS04VYG Tetraodon
31	456.5	18.1	792	13	B1456015	B1456015 603170460
32	456	18.1	420	12	BE929813	BE929813 RC3-GN004
33	455.5	18.1	896	14	BQ955616	BQ955616 AGENCOURT
34	453.5	18.0	877	12	BG681943	BG681943 602629884
35	452	17.9	688	13	B1249235	B1249235 602995511
36	452	17.9	3951	11	BC013626	BC013626 Mus muscu
37	451.5	17.9	921	14	BQ921102	BQ921102 AGENCOURT
38	451.5	17.9	1192	14	BM905999	BM905999 AGENCOURT
39	449.5	17.8	907	14	BQ222937	BQ222937 AGENCOURT
40	448	17.8	946	14	BQ942234	BQ942234 AGENCOURT
41	447	17.7	862	12	BG171919	BG171919 602333020
42	446.5	17.7	608	13	BM425818	BM425818 pgf2c.pk0
43	446	17.7	651	10	BB613976	BB613976 BB613976
44	446	17.7	819	12	BG679800	BG679800 602626481
45	446	17.7	1120	14	BM809933	BM809933 AGENCOURT

ALIGNMENTS

RESULT 1
BC009162
LOCUS
DEFINITION BC009162 Mus musculus, 3305 bp mRNA linear HTC 07-AUG-2002
clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs@remai.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK plate: 5 Row: h Column: 10

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis

This clone has the following problem: incomplete processing.

Location/Qualifiers

1. 3305

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="C57BL/6J"

/clone="IMAGE:2811487"

/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months

old, gross tissue."

/clone_lib="NCL_CGAP_Mam5"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

BASE COUNT 1031 a 757 c 760 g 757 t

ORIGIN

Alignment Scores:

Pred. No.: 6,95e-150 Length: 3305
Score: 2344.00 Matches: 417
Percent Similarity: 96.50% Conservative: 24
Best Local Similarity: 91.25% Mismatches: 16
Query Match: 93.02% Indels: 0
DB: 11 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x BC009162 (1-3305)

Qy 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 951 ATGAATAGCCAGCTCAGCTCATTACAGAGTGCAGATGGACAACATACCACTATCTCACAG 1010
Qy 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 1011 GCCAACGAGCAGAGCCTGAAGACCTTCAGGACTTACACAAGGATACACAAAAATAGAACA 1070
Qy 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 1071 GCTGTCAAGTTCAGCAACTTCAGGAACGCTTCCAGGCTTTTCAGACAGATATGTGAAC 1130
Qy 61 IleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
Db 1131 ATCATTTAGCAACATCAGCTACACAGCCCATCATCTGCGGACACTGACCAAGCAATCTGAAT 1190
Qy 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 1191 GATGTTAGGACCATGTCAGACACACCTTGACACAGACACACGATGACCTGCTCCTTTG 1250
Qy 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnAspLeu 120
Db 1251 AATAACACACTAGTCAACATCGCTGGATTCTATTTCTCTCAGGATGACGACGACATG 1310

RESULT 2

BQ955927

LOCUS

DEFINITION

BQ955927 957 bp mRNA linear EST 21-AUG-2002
AGENCY: NCBI
DEFINITION: AGENCY: NCBI
DEFINITION: AGENCY: NCBI

5', mRNA sequence.
BQ955927
VERSION BQ955927.1 GI:22371405
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cdNA Library Preparation: ResGen, Invitrogen Corp
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13740 row: 1 column: 13
High quality sequence stop: 665.

FEATURES
source Location/Qualifiers
1..957
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6313500"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Olfactory epithelium; Vector: pcwv-SPORF6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: oligo dt. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 260 a 269 c 258 g 169 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 3,26e-90 Length: 957
Score: 1460.50 Matches: 284
Percent Similarity: 93.73% Conservative: 15
Best Local Similarity: 89.03% Mismatches: 13
Query Match: 57.96% Indels: 8
DB: 14 Gaps: 3

US-09-763-712A-2_COPY_91_547 (1-457) x BQ955927 (1-957)

Qy	2	AnsSerGln----	LeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln	20
Db	8	AATTCGGGGATC	TCTCCAGGTACATTACAGAAATTTACCACTATCTCACAG	67
Qy	21	AlaAsnGluGlnAsnLeuLysAspLeuGluAspLeuHisLysAspAlaGluAsnArgThr	40	
Db	68	GCCAACGACGACGCCGTAAGACCTTCAGACATTCACAAGGATACAGAATAAGAAC	127	
Qy	41	AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn	60	
Db	128	GCTGCAAGTTACGCCCAACTTGAGAACGCTTCCAGGCTTTTGAGACAGATATTGTGAAC	187	
Qy	61	IleIleSerAsnIleSerTyThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn	80	
Db	188	ATCATTAGCAACATCAGCTACACAGCCCATCACCTGAGGACACTGACGACCAATCTGAAT	247	
Qy	81	GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu	100	
Db	248	GATGTTAGGACCATGCACAGACACCTTGACACACACAGCGATGACCTGACCTCCTTG	307	
Qy	101	AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu	120	
Db	308	AATAACACACTAGTCAACATCCGCTTGATTTCTATTCTCTCAGAGTGCACAGACATG	367	
Qy	121	MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys	140	

/note="Organ: olfactory epithelium; Vector: pcMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library." 147 t 1 others

BASE COUNT 221 a 253 c 239 g 147 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2,52e-76 Length: 861
Score: 1256.00 Matches: 229
Percent Similarity: 91.01% Conservativeness: 14
Best Local Similarity: 85.77% Mismatches: 20
Query Match: 49.84% Indels: 4
DB: 14 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BQ713873 (1-861)

QY 120 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 139
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 13 ATGATGAGGTCAAGGTAGACACTGAAGTGGCCAACTTATCAGTGGTTATGAAGAGATG 72

QY 140 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhetThrIleLeuGlnGlyPro 159
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 73 AAATGGTGGTCTCAAGCAGCGTCAAGTCTCAAGAACTTTTACCATTCTCAAGGTCT 132

QY 160 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGly 179
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 133 CTGGCCCCAGAGTCCAAAGAGGTACAGAGATCTCAGGACCACCTGGTCCAACTGGC 192

QY 180 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArg 199
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 193 AACAGGGACAGAAAGGAGAGAGAGCGCTGGTCCACCTGGCCCTCGGGGTGAGAGG 252

QY 200 GlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyGlySerLysGlySer 219
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 253 GGCACAAATTTGGACCTGCGCCCTCTGGAGAGCGTGGCAGCAAGGATCCAAAGGCTCA 312

QY 220 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGly 239
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 313 CAGGGTCCCAAGGATCTCGTGGTCCCAAGGAGGAGGCTGGCCCTCAAGGACCTAGTGG 372

QY 240 AspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProPro 259
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 373 GACCCAGGACCCAGGTCACAGGCAAGGATGGACTCCCTGGCCCTCAGGCGCCCTCT 432

QY 260 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 279
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 433 GGGTCCAGGGACTACAGGGCACTGTGGGTGAGCTGGAGTACCTGGACCTCGGGGGTTG 492

QY 280 ProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGly 299
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 493 CCAGGCTTCCAGGGGTGCCAGGATCCCTGGCCCTAAGGGACCACTGGCCCTCCAGGC 552

QY 300 ProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAsp 319
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 553 CCCTCAGGAGCAATGGAGCATTTGGCTCTGCATAATGAACCAACCCCAACATCAGAGTGC 612

QY 320 AsnGlyCysProProHisTrpLysAsnPhetThrAspLysCysTyrThrPheSerVal -G 339
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 613 ACAGGATGTCCGCTCCTGGAAGAACTTTCAGAGATAAATGCTACTATTTTTCCTTGGCA 672

QY 339 uLysGluIlePheGluAspAlaLysLeu -PheCysGluAspLysSerHisLeuValP 359
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 673 AAAAAAATTTTCGAAAAATGCTACCTTTTCTGCGAACAATAATCTCCCCATCCGGGT 732

QY 359 heIleasnThrArgGluGlnGlnTrp -Ile -LysLysGlnMetValGlyArgGluSe 378
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 733 TCATACCCCCCAAGAAAAACAGCATGGGATAAAAAACCTTACCTTCGGGGGACCAAGAAC 792

QY 378 rHisTrpIleGly 382
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 793 CCGCTGGATCGGG 805

RESULT 4
BI456109
LOCUS
DEFINITION
603172765F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5251888 5',
mRNA sequence.
BI456109
BI456109.1 GI:15246765
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 906)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM11636 row: b column: 17
High quality sequence stop: 730.
FEATURES
source
1..906
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5251888"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pcMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 247 a 228 c 256 g 175 t
ORIGIN
Alignment Scores:
Pred. No.: 5,8e-76 Length: 906
Score: 1251.00 Matches: 239
Percent Similarity: 89.12% Conservativeness: 15
Best Local Similarity: 83.86% Mismatches: 27
Query Match: 49.64% Indels: 6
DB: 13 Gaps: 2

US-09-763-712A-2_COPY_91_547 (1-457) x BI456109 (1-906)

QY 76 ThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAsp 95
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 3 ACCGCAATCTCAATGATGTTAGGACCACATGCACAGACACCTTGACACACACGAT 62

QY 96 AspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleAerGluAspSerValSerLeuArg 115
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 63 GACCTGACCTCTTGAATAACACACTAGTCAACATCCGCTGGATTTCTTCTCTCAGG 122

QY 116 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 135
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 123 ATGCAGCAGACATGATGAGTCAAGTTAGACTGGAAGTGGCCCACTTATCAGTGGT 182

QY 136 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhetThrIle 155
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 183 ATGAAGAGATGAACCTGGTTCCTCAAGCAGCGTCACTCATCAAGAACTTTACCAT 242

QY 156 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 175


```

|||||
Db 243 CTACAGGTCTCCCTGGGCCAGAGGTCCAAAGGTGACAGAGGATCTCAGGAGCACCT 302
QY 176 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 195
Db 303 GTCCACTGGCAACAGGACAGAAAGAGAGAGGAGAGCCCTGGTCCACTGGCCCT 362
QY 196 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGly 215
Db 363 GCGGTGAGAGGGGACAAATTTGACCAAGTCAGTCGCCCTCTCTGGAGAGCGTGGCAGCAAGGA 422
QY 216 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 235
Db 423 TCCAAAGGCTACAGGGTCCCAAGGATCTCGTGGGTCCCCAGTGAAGCCTGGCCCTCAA 482
QY 236 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 255
Db 483 GGACTAGTGGGGACCCAGGACACAGTCACAGGCAAGATGGATCTCCCTGGCCCT 542
QY 256 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 275
Db 543 CAGGCGCTCTCTGGCTTCCAGGACTACAGGCGACTGTGGGTGAGCTGGAGTACCTGGA 602
QY 276 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 295
Db 603 CCTCGGGGTTCACAGGCTTGCAGGGGTGCGAGGATGCCCTGGGCTTAAGGAGCACT- 661
QY 296 GlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 315
Db 662 AGCCCTCAGGCGCCCTCAGGACATG-GAGCCATGGCTGCAGAAATGAACACAGCCCA 720
QY 316 AlaProGlu--AspAsnGlyCysProProHisTrpLysAsnPheThrAsp---LysCysT 334
Db 721 GCATCAGAGGTTCAACGCGATGTCCTCTCACTCGAAGAACTTTCACGGGATTAATGGCT 780
QY 334 yTyrPheSerVal---GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspL 353
Db 781 ACTATTTTCCATTCACAAAAGAAATTTTGAAGATTTCTGAAGCGTTCTGTGGAGAA 840
QY 353 ySerSerHis 356
Db 841 AATTTTCCAAT 851

RESULT 5
BQ891432
LOCUS
DEFINITION
AGENCOURT_8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
5', mRNA sequence.
ACCESSION
BQ891432
VERSION
BQ891432.1
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 936)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13790 row: n column: 19
High quality sequence start: 25
High quality sequence stop: 632.
Location/Qualifiers

```

```

source
1..936
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6332754"
/lab_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb;
Site_1: EcorV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 273 a 250 c 244 g 163 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 3,64e-75 Length: 936
Score: 1239.50 Matches: 241
Percent Similarity: 94.49% Conservative: 16
Best local Similarity: 88.60% Mismatches: 9
Query Match: 49.19% Indels: 6
DB: 14 Gaps: 1
US-09-763-712a-2_COPY_91_547 (1-457) x BQ891432 (1-936)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrLeSerGln 20
Db 131 ATGAATAGCCAGCTCAGCTCATTCACAGTCAGATGGCAACATTACCACTATCTCACAG 190
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 191 GCCAAGCAGCAGGCTGAAGACCTTCAGGACTTACACAGGATACAGAAATAGAAC 250
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 251 GCTGTCAAGTTTCAGCCAACTTCAGGAACGCTTCCAGGCTTTTCAGACAGATATTGTGAAC 310
QY 61 IleLeSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
Db 311 ATCATTAGCAACATCAGCTACACAGCCCATCAGCTGAGGACATGACCACTGCAAT 370
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 371 GATGTTAGACCACTGACAGACACCTTGACCAAGACACAGGATGACCTGACCTCTTG 430
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 431 ATAACACACTAGTCAACATCCCTGGATTTCTTCAGGATGACCAAGACATG 490
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
Db 491 ATGAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCAGTGGTTATGAAGAGATGAA 550
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
Db 551 CTGGTTGACTCAAGCAGCGTCAGCTCACTCAAGAACTTTACCACTTCAAGGTCCTCCT 610
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 180
Db 611 GSCCCAGAGGTCCAAAAGAGTGACAGAGATCTCAGGACCACTGCTGCTCAACTGGCAAC 670
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
Db 671 AAGGGACAGAAAGGAGAGAGAGGAGAGCGCTGGTCCACCTGGCCCTGCGGGTGAGAGGGG 730
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySerGln 220
Db 731 ACAATTTGGACCACTCGGCCCTCTGAGAGAGCGTGGCAGCAAGGATCCAAAGGCTCACAG 790
QY 221 Gly-ProLysGlySerArgGlySer--ProGlyLysProGlyProGlnGlyProSerGlyA 240
Db 791 GGNATCAAGGGATCTCGTGGGTGCCCGGAGGAGGCTGGCCCTCAAGGACTAGTGGGG 850
QY 240 spProGly-ProProGlyProProGly-LysGluGlyLeuPro-GlyProGlnGlyProP

```

```

||||| 851 ACCAGNACCCAGGTCACAGGNCAGGATGAGTACCTCGGCGCCCTCAGGNCNT 910
||||| 259 ro---GlyPheGlnGlyLeuGlnGly 266
||||| 911 CCTGGGCTCNCAGGGGACTACAGGG 936

RESULT 6
BE910803 580 bp mRNA linear EST 29-SEP-2000
LOCUS 601661855F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962292 5',
DEFINITION mRNA sequence.
ACCESSION BE910803
VERSION BE910803.1 GI:10407765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LL9128 row: m column: 13
High quality sequence stop: 580.
FEATURES
Location/Qualifiers
1..580
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3962292"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
Note="Organ: Mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 159 a 153 c 161 g 107 t
ORIGIN
Alignment Scores:
Pred. No.: 1..5e-60 Length: 580
Score: 1022.00 Matches: 177
Percent Similarity: 96.34% Conservative: 7
Best Local Similarity: 92.67% Mismatches: 7
Query Match: 40.56% Indels: 0
DB: 12 Gaps: 0
US-09-763-712a-2_COPY_91_547 (1-457) x BE910803 (1-580)

QY 214 LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGly 233
||||| 6 AAAGATCAAAAGGCTCACAGGTCCTCCAAAGGATCTCGTGGTCCCGAGGAAAGCTGGC 65
QY 234 ProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlnGlyLeuPro 253
||||| 66 CCAAGGACCTAGTGTGGGACCCAGGACCAAGGTCACAGGCAAGGATGACATCCCT 125
QY 254 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
||||| 126 GGCCCTCAGGGCCCTCTGGCTCCAGGACTACAGGGGACATGTGGTGGGCTGGAGTA 185

```

```

QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 186 CCTGGACCTCGGGGTTGCCAGGCTTCCAGGGGTGCCAGGATGCTGGCGCTTAAGGA 245
QY 294 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
Db 246 CCACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCCATTTGGCTCTGCAGATGAACCA 305
QY 314 ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCys 333
Db 306 ACCCAGCATCAGAGTCAACGGATGTCGCCCTCACTGGAAGAACTTCACAGATAATGC 365
QY 334 TyrTyrPheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLys 353
Db 366 TACTATTTTTCATTTGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAA 425
QY 354 SerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMet 373
Db 426 TCTTCCCATCTCGTTTTCATAAATCAAGAGAAAGACACGAATGGATAAAAAAGCATACC 485
QY 374 ValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 393
Db 486 GTGGGAGAGAAAGCCATTTGGATCGGCCCTCACAGACTCAGACAGGAAGCAATGGAAG 545
QY 394 TrpLeuAspGlyThrSerProAspTyrLysAsn 404
Db 546 TGGCTAGAGGGGTACCTGTTGATTACAAAAA 578

RESULT 7
BQ674807 668 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672
DEFINITION 5', mRNA sequence.
ACCESSION BQ674807
VERSION BQ674807.1 GI:21785641
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM2377 row: c column: 09
High quality sequence stop: 656.
FEATURES
Location/Qualifiers
1..668
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6212672"
/clone_lib="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
Note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 216 a 148 c 167 g 137 t
ORIGIN

```

Alignment Scores:	1.82e-59	Length:	668
Pred. No.:	Score:	Matches:	174
	1007.00	Conservative:	0
Percent Similarity:	99.43%	Mismatches:	1
Best Local Similarity:	99.43%	Indels:	0
Query Match:	39.96%	Gaps:	0
DB:	14		

US-09-763-712A-2_COPY_91_547 (1-457) x BQ674807 (1-668)

283	ProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyProSerGly	302
QY		
3	CCTGGGTACACAGGCATGCCAGGCCCAAGGCCGCCCGGCCCTCTGGCCCATCAGGA	62
DB		
303	AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCys	322
QY		
63	GCGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGGCCACCGAGGACCAATAGCTGC	122
DB		
323	ProProHisTrpLysAsnPheThrAspLysCysTyTrpPheSerValGluLysGluIle	342
QY		
123	CCGCCTCACTGGAGAACTTCACAGCAAAATGCTACTATTTTTTCAGTTGAGAAAGAAAT	182
DB		
343	PheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThr	362
QY		
183	TTTGGAGTGTCAAAAGCTTTTCTGTGAAGACAAGCTTTCACATCTTGTTTTCATAACT	242
DB		
363	ArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGly	382
QY		
243	AGAGAGGAACAGCAATGGATAAAAAACAAGATGGTAGGGAGAGAGACCGCCTGGATCGGC	302
DB		
383	LeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyr	402
QY		
303	CTCACAGACTCAGAGCGTGAATAAGTGAAGTAGTGCTGGATGGACACTCTCCAGACTAC	362
DB		
403	LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAsp	422
QY		
363	AAAAATTGGAAGCTGGACAGCCGGATACTGGGGTCATGGCCATGGGCCGAGGAAGAC	422
DB		
423	CysAlaClyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsn	442
QY		
423	TGTGCTGGGTTGATTTATCTGGCCAGTGGACGATTTCCTCAATGTGAAGACGCTCAATAC	482
DB		
443	PheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu	457
QY		
483	TTCAATTGGCAAAAGACAGGAGACAGTACTGTCTATCTGCATTA	527
DB		

RESULT 8	
BE290299	
LOCUS	601 bp mRNA linear EST 26-OCT-2000
DEFINITION	601089246f1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5',
ACCESSION	BE290299
VERSION	BE290299.1
KEYWORDS	GI:9171250
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 601)	NIH-MGC http://mgc.mci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys			
	CDNA Library Preparation: Life Technologies, Inc.			
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)			
	Clone distribution: Incyte Genomics, Inc.			
	found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
	plate: LLAM8516	row: k	column: 12	

High quality sequence stop: 551.
Location/Qualifiers
1. .601

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3484163"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stages="7 months"
/lab_host="DH10 $\alpha$ "

```

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

N11			
BASE COUNT	156 a	161 c	179 g 105 t
ORIGIN			

Alignment Scores:	4.18e-56	Length:	601
Pred. No.:	957.00	Matches:	183
Score:	93.10%	Conservative:	6
Percent Similarity:	90.15%	Mismatches:	11
Best Local Similarity:	37.98%	Indels:	6
Query Match:	10	Gaps:	0
DB:			

US-09-763-712A-2_COPY_91_547 (1-457) x BE290299 (1-601)

[illegible]

RESULT 9

AL568743/c
 LOCUS
 DEFINITION AL568743 LTI_FL002_PL1 Homo sapiens CDNA clone CS0DE005YH04 3 prime
 mRNA sequence.
 ACCESSION AL568743
 VERSION AL568743.1 GI:129233387
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 967)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 1..967
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DE005YH04"
 /clone_lib="LTI_FL002_PL1"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 222 a 221 c 217 g 290 t 17 others
 ORIGIN

Alignment Scores:
 Pred. No.: 3-82e-53 Length: 967
 Score: 916.50 Matches: 174
 Percent Similarity: 91.1% Conservative: 1
 Best Local Similarity: 90.62% Mismatches: 14
 Query Match: 36.37% Indels: 5
 DB: 9 Gaps: 3

US-09-763-712a-2_COPY_91_547 (1-457) x AL568743 (1-967)

QY 266 GlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyVal 285
 |||||
 DB 947 GGCACCGTTGGGAGCGTGGTCT---GGACCTCGGGAGCTGCCAG--CTTGCCTGGGTA 894
 QY 286 ProGlyMetProGlyProGlyProGlyProGlyProGlyProGlyProSerGlyAlaValVal 305
 |||||
 DB 893 CCAGGATCGCAGCCCAAG---CCCGCCGCCCTCTCTGC-CCATCAGGAGCGGTGTG 838
 QY 306 ProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProProHis 325
 |||||
 DB 837 CCCTGSCCTCGAGAAATGAGCAACCCCGGACCCGGAGGACAAATGGCTGCCGCTCAC 778
 QY 326 TrpLysAsnPhetrAspLysCysTyrTyrrPheSerValGluLysGluIlePheGluAsp 345
 |||||
 DB 777 TGGAAGAACTTCACACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGAT 718
 QY 346 AlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlu 365
 |||||
 DB 717 GCAAGCTTTCTGTGAAGCAAGTCTTCACATCTTGTGTTTTCATAAACACTAGAGAGAA 658
 QY 366 GlnGlnTrpIleLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAsp 385
 |||||
 DB 657 CAGCAATGATATAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCTCACAGAC 598

QY 386 SerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrIlysAsnTrp 405
 |||||
 DB 597 TCAGACCGTGAAATCAATGGAAGTGGCTGGATGGACATCTCCAGACTACAAAAATGG 538
 QY 406 LysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 425
 |||||
 DB 537 AAAGCTGGACAGCCGATAACTGGGTCATGC-CATGGGCCAGGAGAACTGTGCTGG 479
 QY 426 LeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCys 445
 |||||
 DB 478 TTGATTATGCTGGGAGTGGAGACGATTTCCATGTGAAGACGTCATTAACCTTCATTGC 419
 QY 446 GluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 DB 418 GAAAAAGACAGGAGAGACAGYACTGTCTATCTGCATTA 383
 RESULT 10
 AW958053
 LOCUS
 DEFINITION EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW958053
 VERSION AW958053.1 GI:8147736
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
 , I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
 Quackenbush, J.
 TITLE Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 JOURNAL Unpublished (2000)
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 115
 Seq primer: Reverse.
 Location/Qualifiers
 1..552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGE"
 /note="Vector: pBluescriptSKm"
 BASE COUNT 176 a 112 c 143 g 121 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.63e-48 Length: 552
 Score: 842.00 Matches: 149
 Percent Similarity: 98.68% Conservative: 1
 Best Local Similarity: 98.03% Mismatches: 1
 Query Match: 33.41% Indels: 1
 DB: 10 Gaps: 0
 US-09-763-712a-2_COPY_91_547 (1-457) x AW958053 (1-552)
 QY 307 LeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrp 326
 |||||
 DB 1 CTGGCCCTCGAATGAGCAACCCCGGACCCGGAGGACAAATGGCTGCCCTCACTGG 60
 QY 327 LysAsnPhetrAspLysCysTyrTyrrPheSerValGluLysGluIlePheGluAspAla 346
 |||||
 DB 61 AAGAAGCTTCACACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCA 120
 QY 347 LysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlu 366
 |||||
 DB 121 AAGCTTTTCTGTGAAGCAAGTCTTCACATCTTGTGTTTTCATAAACACTAGAGAGAACAG 180

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

source

Location/Qualifiers
 1. .638
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A730023E20"
 /clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum"
 /tissue_type="cerebellum"
 /dev_stage="7 days neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATAATTAATCCCTCCCTCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 212 a 134 c 146 g 146 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.7e-47 Length: 638
 Score: 831.00 Matches: 142
 Percent Similarity: 93.25% Conservative: 10
 Best Local Similarity: 87.12% Mismatches: 11
 Query Match: 32.98% Indels: 0
 DB: 10 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BB248064 (1-638)

Qy 295 ProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThr 314
 |||||
 Db 3 CCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCATGGCTCGCATGAACCAACC 62
 Qy 315 ProAlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTyr 334
 |||||
 Db 63 CCAGCATAGAGGATCAACGGATGTCGCCCTCACTGGAAGAACTTCACAGATAAATGCTAC 122
 Qy 335 TyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 354
 |||||
 Db 123 TATTTTTCATTGGAAAAAGAAATTTTGAAGATGCTTAAGCTTTCTGTGAAGCAAAATCT 182
 Qy 355 SerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetVal 374
 |||||
 Db 183 TCCCATCTCGTTTTCATAACTCAAGAGAGAACACGATGGATAAAAAAGCATACCGTG 242
 Qy 375 GlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrp 394
 |||||
 Db 243 GGGAGAGAAAGCCATTGGATCGCCCTCACAGACTCAGACAGAGAAAGCGAATGGAAGTGG 302
 Qy 395 LeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGly 414
 |||||
 Db 303 CTAGACGGGTCCCTCTTGTATTACAAAACTGGAAAAAGCTGGCAACACAGATAACTGGGGC 362

Qy 367 GlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSer 386
 |||||
 Db 181 CAATGGATAAAAAACAGATGCTAGGAGAGAGGCCACTGGATCGGCTCACAGACTCA 240
 Qy 387 GluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLys 406
 |||||
 Db 241 GAGCGTGAATAATGAATGGAATGGTGGATGGACATCTCCAGACTACAAAAATGGAAA 300
 Qy 407 AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 426
 |||||
 Db 301 GCTGGACAGCCGGATAACTGGGGTGTATGCCATGGCCAGGAGAGACTGTCTGGGTG 360
 Qy 427 IleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhe-IleCysG1 446
 |||||
 Db 361 ATTTATGTGGCGCAGGGAACGATTTCCTCAATGTGAGAGCTCAATAACTTTAATTGGCA 420
 Qy 446 uLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 421 AAAAGACAGGAGAGAGTACTGGCATCTGCATTA 454

RESULT 11

BB248064

LOCUS

DEFINITION

BB248064 638 bp mRNA linear EST 23-OCT-2001
 musculus cDNA clone A730023E20 3', mRNA sequence.

ACCESSION

BB248064

VERSION

BB248064.2

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 638)

Arkawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,Y., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
 ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
 ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Jul 6, 2000 this sequence version replaced gi:8940810.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,K.I., Aizawa

,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.

```

QY 415 HisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnThrPasnAsp 434
Db 363 AGTGGCATGGGCCAGAGAAAGACTGTGGCTTATTTACCGACAGCAGTGAATGAC 422
QY 435 PheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlnThrValLeuSer 454
Db 423 TTCCAGTGTGATGAATCAATAACTTCAATTTGTGAGAGAAAGAGGAGGAGTACCATCA 482
QY 455 SerAlaLeu 457
Db 483 TCCATATTA 491

RESULT 12
BQ934501 884 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8765685 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330794
DEFINITION 5', mRNA sequence.
ACCESSION BQ934501
VERSION BQ934501.1 GI:22349884
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM13785 row: m column: 03
High quality sequence stop: 352.
FEATURES
    source
        1..884
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:6330794"
            /clone_lib="NIH_MGC_130"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb; Site_1: ECKRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 282 a 225 c 205 g 172 t
ORIGIN

Alignment Scores:
Pred. No.: 3,84e-45 Length: 884
Score: 798.50 Matches: 170
Percent Similarity: 90.24% Conservative: 15
Best Local Similarity: 82.93% Mismatches: 12
Query Match: 31.69% Indels: 9
DB: 14 Gaps: 4

US-09-763-712A-2_COPY_91_547 (1-457) x BQ934501 (1-884)

QY 1 MetAsnSerGlnLeuAsnSerPheThr-GlyGlnMetGluAsnIleThrThrIleSerG1 20
Db 277 ATGAATAGCAGCTCAGCTCATTACAGAGGTGAGTGGACAACATTACCACTATCTACA 336
QY 20 nAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgTh 40
Db 337 GGCACACGACGAGAGCTCAAGACCTTCAGGACTTACACAGGATACAGAAATAGAAC 396

```

```

QY 40 rAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValas 60
Db 397 AGCTGTCAAGTTTCAGCAACTTCAGGAACGCTTCCAGGTCTTTGAGACAGATATTGTGAA 456
QY 60 nIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAs 80
Db 457 CATCATTAGCAACATCAGCTACACGCCCATCACCTGAGGACACTGACCAGCAATCTGAA 516
QY 80 nGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerIe 100
Db 517 TGATGTTAGGACCACATGCACACACACTTGACCAGACACACCGATGACCTGACCTCCTT 576
QY 100 uAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLe 120
Db 577 GAATAC-ACACTAGTCAACATCCGCTTGGATTCTATTTCTCAGGATGAGCAAGACAT 635
QY 120 uMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLy 140
Db 636 GATGAGGTCAAGTTAGACACTGAAGTGCCAACTTATCAGTGGTTATGGAAGAGATGAA 695
QY 140 sLeuValAspSerLysHisGlyGlnLeuLysAsnPheThrIleLeuGlnGlyProPr 160
Db 696 ACTGGTTGACTCAACACCGGTCAGCTCATCAAGAACTTTACCTTCTACAGGTCTCTCC 755
QY 160 oGlyProArgGly-ProArgGlyAspArgGlySerGlnGlyProPro---GlyProThr- 178
Db 756 TGGCCCCAGAGGTCAAAAAGGGGGGACAGAGGATCTCAGGGGACCAACTGGGTCCAACAT 815
QY 179 GlyAsnLysGlyGlnLys-----GlyGluLysGlyGluPro---GlyProPro---Gly 194
Db 816 GGCACAAAGGGGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
QY 195 ProAlaGly 197
Db 876 CCTGGCGGC 884

RESULT 13
BQ771366
LOCUS DEFINITION 808 bp mRNA linear EST 26-JUL-2002
ACCESSION BQ771366
VERSION BQ771366.1 GI:21979842
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
FEATURES
    source
        Location/Qualifiers
            1..808
                /organism="Mus musculus"
                /strain="C57BL/6"
                /db_xref="taxon:10090"
                /clone="IMAGE:5702432"
                /clone_lib="NIH_BMAP_F10"
                /tissue_type="whole brain"
                /dev_stage="embryo 12.5dpc"

```


QY 380 TrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSer 399
 |||||
 Db 375 TGGATCGGCTCACAGACTCAGACGCTGAAATGAATGAAAGTGGCTGGATGGACATCT 316
 |||||
 QY 400 ProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyPro 419
 |||||
 Db 315 CCAGACTACAAATTTGGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGGCCA 256
 |||||
 QY 420 GlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAsp 439
 |||||
 Db 255 GGAGAGACTGCTCGGGTGTATTATGCTGGCAGTGGAGCAATTCCAATGTGAGAC 196
 |||||
 QY 440 ValAsnAspPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 195 GTCAATAACTTCAATTTGCCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 142
 |||||

RESULT 15
 BM713891

LOCUS
 DEFINITION UI-E-EJ0-ahq-c-05-0-UI-r1 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ahq-c-05-0-UI 5', mRNA sequence.

ACCESSION BM713891
 VERSION
 KEYWORDS EST.
 SOURCE BM713891.1 GI:19027149
 human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

9704477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue procurement: Dr. Gregg Hageman
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA
 sequence: 443-498, >Poly_A/simple_repeat
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..507
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ahq-c-05-0-UI"
 /clone_lib="UI-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (life technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAAGA

; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
 optic nerve, CCATTAAAGT; retina, CCGCG; Retina Foveal and
 Macular, GTCC; RPE and Choroid, ACGTA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT 184 a 89 c 121 g 112 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 3.81e-36 Length: 507
 Score: 663.00 Matches: 116
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 26.31% Indels: 0
 DB: 14 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x BM713891 (1-507)

QY 342 IlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsn 361
 |||||
 Db 3 ATTTTGTAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATAAAC 62
 |||||
 QY 362 ThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIle 381
 |||||
 Db 63 ACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATC 122
 |||||
 QY 382 GlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAsp 401
 |||||
 Db 123 GGCCTCAGACTCAGACGGTGGAATAAATGAATGGAAGTGGCTGGATGGACATCTCCAGAC 182
 |||||
 QY 402 TyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGlu 421
 |||||
 Db 183 TACAAAAATTTGGAAGCTGGACAGCCGGATAAATCTGGGTCTATGGCCATGGGCCAGAGAA 242
 |||||
 QY 422 AspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsn 441
 |||||
 Db 243 GACTGTCTGGGTGATTATGCTGGCAGTGGAAACGATTTCCAATGTGAAGACGTCAT 302
 |||||
 QY 442 AsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 303 AACTTCATTTGGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 350
 |||||

Search completed: March 21, 2003, 12:20:27
 Job time : 1517.54 secs

FEATURES source

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 06:54:47 ; Search time 3021.43 Seconds
(without alignments)
4401.885 Million cell updates/sec

Title: US-09-763-712a-2_COPY_91_547

Perfect score: 2520

Sequence: 1 MNSQLNSFTGQENITITISQ.....EDVNNFICEKDKRTVLSSAL 457

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09763712/runat_14032003_100949_18101/app_query.fasta_1.1877
-DB=genEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09763712 -ACGN_1_1.6828 @runat_14032003_100949_18101 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -NLOGLG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: gb_wa.*
16: em_ba.*
17: em_fun.*
18: em_in.*
19: em_mu.*
20: em_mu.*
21: em_om.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

RESULT 1

ALIGNMENTS

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	2520	100.0	2983	9	AB005145 Homo sapi
2	2514	99.8	2641	6	AX047353 Sequence
3	2514	99.8	3058	9	AB038518 Homo sapi
4	2487	98.7	2005	6	AX454442 Sequence
5	2487	98.7	2005	6	AX490920 Sequence
6	2344	93.0	2637	10	AB078434 Mus muscu
7	2330	92.5	3291	10	AB038519 Mus muscu
8	2309	91.6	1886	9	HSB02985
9	2244	89.0	4330	9	AB052103
10	1021.5	40.5	130763	2	AC112416 Rattus no
11	981	38.9	182029	2	AC114677 Mus muscu
12	981	38.9	193208	2	AC102618 Mus muscu
13	954.5	37.9	169088	2	AC016128 Homo sapi
14	954.5	37.9	188439	9	AF000915 Homo sapi
15	935.5	37.1	71044	2	AC024368 Homo sapi
16	916.5	36.4	187635	2	AF001022 Homo sapi
17	914.5	36.3	178022	2	AF000900 Homo sapi
18	842	33.4	188255	2	AF000939 Homo sapi
19	763	30.3	188255	2	AF000939 Homo sapi
20	747	29.6	193208	2	AC114677 Mus muscu
21	747	29.6	193208	2	AC102618 Mus muscu
22	682	27.1	3636	9	AB007829 Homo sapi
23	682	27.1	3685	6	E32511 Scavenger r
24	682	27.1	3810	6	E32509 Scavenger r
25	671	26.6	2215	10	BC026446 Mus muscu
26	562	22.3	188637	2	AL845306 Danio rer
27	491.5	19.5	1265	10	RATSPD M81231 Rat pulmona
28	483.5	19.2	3508	3	AF053538 Alvinella
29	481.5	19.1	1183	10	BC003705 Mus muscu
30	481.5	19.1	1253	10	MUSSPD L40156 Mus muscu
31	474.5	18.8	1410	6	AX334792 Sequence
32	474.5	18.8	1410	9	HSMRNAPD X65018 H.sapiens m
33	474	18.8	1454	4	BTLSPD X75911 B.taurus mR
34	473.5	18.8	1284	4	AF509590 Bos tauru
35	473.5	18.8	1301	9	BC022318 Homo sapi
36	471	18.7	2377	9	BC008760 Homo sapi
37	471	18.7	5676	6	E07265 D90279 Human mRNA
38	471	18.7	5676	9	HUMPAIV M76729 Human pro-a
39	471	18.7	7138	9	HUMPAIV M76730 Chinese ham
40	465	18.5	6114	10	CRUPAIV L06863 Cricetulus
41	462.5	18.4	4990	10	CRUCOLVIA U07973 Gallus gall
42	462	18.3	2777	5	GG007973 AX146422 Sequence
43	462	18.3	4428	6	AX146424 Sequence
44	462	18.3	4428	6	AX146424 Sequence
45	462	18.3	5575	5	AF137273 Gallus ga

```
AB005145
LOCUS       AB005145             2983 bp    mRNA    linear    PRI 21-NOV-2001
DEFINITION Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION   AB005145
VERSION     AB005145.1  GI:17026100
KEYWORDS    .
SOURCE      Homo sapiens female tissue_lib:placenta cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
            Sakai,Y., Fukuh,A., Sakamoto,T., Itabe,H., Suzutani,T.,
            Ogasawara,M., Yoshida,I. and Wakamiya,N.
            The membrane-type collectin CL-P1 is a scavenger receptor on
            vascular endothelial cells
            J. Biol. Chem. 276 (47), 44222-44228 (2001)
JOURNAL     21570232
MEDLINE     2 (bases 1 to 2983)
REFERENCE   Ohtani,K.
AUTHORS     Direct Submission
TITLE       Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
            Department of Microbiology; 2-1-1-1 Midorigaoka-Higashi, Asahikawa,
            Hokkaido 078-8510, Japan (E-mail:ohtani@asahikawa-med.ac.jp,
            Tel:+81-166-68-2393, Fax:+81-166-68-2399)
FEATURES    Location/Qualifiers
            1..2983
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /sex="female"
             /tissue_lib="placenta"
            1..2983
             /gene="CL-P1"
             /gene="CL-P1"
             /codon_start=1
             /product="collectin placenta 1"
             /protein_id="BAB72147.1"
             /db_xref="GI:17026101"
             /translation="MKDDFAEEVEVSFGYKRFQIGTQCTKCKNNWALKPSILLY
            ILCALLTTVAIGYKVKVMDNTVGMETSROTVDKLLTAVESDLKLGDOTGKKAI
            STNSLSTFRSDILDRLQOLREITEKTSKNKDTLEKLAQSGDALVDRSOLKETLENN
            SFLITTVNKTQAYNGYVTNLQDDTSVLQGNLQNMYSHVYVIMLNNLNTFQVQQRN
            LITNQRSVDDTSQAIRIKNDQFNLOQVFLQAKDQTLKQVLSQTLAANNLSALA
            KANDTLMDSLQSFQOMENITTSQANEQNLKDLQDLHKDAENRTAIKFNQLEE
            RFPLETDIVNISISYTAHLRLTSLNLEVRTCTDTLT KHTDDLTSLNLTANI
            RLDSVLRMQDLMRSRLDTEVANLSVIMEKMLVDSKHGOLIKNFTILOGPPGPRGP
            RGDRSQPPGPTGNKQKQKGEKGPFGPAGERGPIGPAGPPGERGKSGSGOPK
            GSRGSPGPGPGSPGDPGPPGKGLPQPGPPGPGFQGLQGTGVEGPGVPGPLPG
            LPVFGPGKPGPPGPGSGAVPALQNEPTPAPEDNGCPPHKNFTDKCYFSVE
            KEIFEDAKLFCEDKSSHLVFINTRREQQWIKQWVGRESHWIGLWDSRENEKWLQD
            TSPDYKNNKAGQPDNMWGHGPGEDCAGLIYAGQWDFQCEDVNNFICEKDRETVLSS
            AL"
BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN
Alignment Scores:
Pred. No.:      2,57e-82      Length:      2983
Score:          2520.00      Matches:     457
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              9           Gaps:         0

US-09-763-712A-2_COPY_91_547 (1-457) x AB005145 (1-2983)
QY      1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
        |||||
DB      926 ATGAACAGCAGCTCAACTCAITTCACAGTCAGATCGAGAACATCACCATTCTCTCAA 985
        |||||
QY      21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
        |||||
DB      986 GCACAGCAGCAACCTGAAGACCTTCAGAGGACCTTACACAAAGATGAGAGATAGAAC 1045
        |||||
```

```
QY      41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
        |||||
DB      1046 GCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 1105
        |||||
QY      61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrIleThrSerAsnLeuAsn 80
        |||||
DB      1106 ATCATTTAGCAATATCATGTTTACACAGCCCACTCCCTCGGACGCTGACCAAGCAATCAAAT 1165
        |||||
QY      81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
        |||||
DB      1166 GAAGTCAGAGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTCTTG 1225
        |||||
QY      101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
        |||||
DB      1226 AATAATACCTCGCCCAACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAAGATTGTG 1285
        |||||
QY      121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
        |||||
DB      1286 ATGAGTCGAGGTTAGACACTGAAGTAGCCAATTTATCAGTGATTATGGAAGAAATGAAG 1345
        |||||
QY      141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
        |||||
DB      1346 CTAGTAGACTCCAAGCATGTCAGCTCATCAAGATTTTACATACTACAAGCTCCACCG 1405
        |||||
QY      161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
        |||||
DB      1406 GGCCCCAGGGGTCCAAGAGGTGACAGAGGATCCAGGGACCCCTGGCCCACTGGCCAAC 1465
        |||||
QY      181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGluArgGly 200
        |||||
DB      1466 AAGGGACAGAAAGAGAGAGGGGAGCGCTGGACACCTGGCCCTCGGGGTGAGAGGCG 1525
        |||||
QY      201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
        |||||
DB      1526 CCATTTGGACCACTGGTCCCCCGGAGAGCGGTGGCGCAAGATCTAAAGCTCCACG 1585
        |||||
QY      221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlyProGlyProSerGlyAsp 240
        |||||
DB      1586 GGCCCCAAAGGGTCCCGTGGTTCCTGGGAAGCCCGGCGCTCAGGGCCCGCAGTGGGGAC 1645
        |||||
QY      241 ProGlyProGlyProGlyProGlyLysGlyGluLeuProGlyProGlyProGlyProGly 260
        |||||
DB      1646 CCAGGCCCCCGGGGCCCAACAGGAGGGGACTCCCCGGGCCCTCAGGGCCCTCTCTGGC 1705
        |||||
QY      261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
        |||||
DB      1706 TTCCAGGGACTTCAGGGCACCGTTGGGAGCCTTGGGTGCCTGGACCTCGGGGACTGCCA 1765
        |||||
QY      281 GlyLeuProGlyValProGlyMetProGlyProGlyProGlyProGlyProGlyPro 300
        |||||
DB      1766 GGCTTGCTGGGTACCAGGCATGCCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGGCCCA 1825
        |||||
QY      301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
        |||||
DB      1826 TCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCCAACCCCGCACGGAGGACAAT 1885
        |||||
QY      321 GlyCysProProHisThrLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
        |||||
DB      1886 GGCTGCCCGCCTCCTGAGAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1945
        |||||
QY      341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
        |||||
DB      1946 GAAATTTTGGAGGATGCAAGGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTTCATA 2005
        |||||
QY      361 AsnThrArgGluGlnGlnThrIleLysLysGlnMetValGlyArgGluSerHisTrp 380
        |||||
DB      2006 AACACTAGAGAGAACACAAATGGATAAAAAACAGATGGTAGGGAGAGAGACCACTGG 2065
        |||||
QY      381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
        |||||
DB      2066 ATCGGCTCACAGACTCAGAGCGGTGAAGATGAATGAATGGGTGGTGGATGGGACATCTCCA 2125
        |||||
```

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
|||||
Db 2126 GACTACAAAATTGGAAAGCTGCACAGCCGATAAAGTGGGTTCATGGCCATGGCCAGGA 2185
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
|||||
Db 2186 GAAGACTGCTGGGTGATGATGCTGGGAGTGAAGCATTTCCAAATGTTGAAGACGTC 2245
QY 441 AsnAspPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
|||||
Db 2246 AATAACTTCATTGGCAAAAGACAGGAGACAGTACTGTCTGCATTA 2296

RESULT 2
AX047353 2641 bp DNA linear PAT 15-DEC-2000
LOCUS
DEFINITION
Sequence 39 from Patent WO0068380.
ACCESSION
AX047353
VERSION
AX047353.1 GI:11876599
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS
Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R.,
Lu,D.A. and Azimzai,Y.
TITLE
Extracellular matrix and adhesion-associated proteins
JOURNAL
Patent: WO 0068380-A 39.16-NOV-2000;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CB1"

BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN
Alignment Scores:
Pred. No.: 3.82e-82 Length: 2641
Score: 2514.00 Matches: 456
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x AX047353 (1-2641)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 567 ATGAACAGCGCAGCTCAACTCACTTACAGGCTCAGATGGAGAACATCACCTATCTCTCAA 626
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisIleAspAlaGluAsnArgThr 40
Db 627 GCCAACAGCAGACAGCTGAAAGACCTTCAGAGACTTACAAAGATGACAGAAATAGAACA 686
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 687 GCCATCAAGTCAACCACTGGAGGAACGCTTCCAGCTCTTGGAGACGATATTGTGAAC 746
QY 61 IleIleSerAsnIleSerThrAlaHisHisLeuAspGluThrLeuThrSerAsnLeuAsn 80
Db 747 ATCATTTAGCAATATCATAGTTACAGACCCACCACTCTCGGACGCTGACCAATCTAAAT 806
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 807 GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 866
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 867 AATAATACCTGGCCAACTCGTTGGATCTGTTCTCTCAGGATGCAACAAGATTG 926
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
|||||

Db 927 ATGAGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG 986
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
|||||
Db 987 CTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAAAGTCCACCG 1046
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
|||||
Db 1047 GGCCCCAGGGTCCAAGAGGTGACAGAGATCCAGGACCCCTGGCCCACTGGCAAC 1106
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
|||||
Db 1107 AAGGGACAGAAAGAGAGAGGGGAGCGCTGGACCACTGGCCCTGCGGTGAGAGAGGC 1166
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyLysGlySerLysGlySerGln 220
|||||
Db 1167 CCAATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCAG 1226
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
|||||
Db 1227 GGCCCAAGGCTCCCGTGGTTCCTGGGAAGCCCGCCCTCAGGGCCCACTGGGGAC 1286
QY 241 ProGlyProProGlyProGlyLysGlyGluLeuProGlyProGlnGlyProGlyPro 260
Db 1287 CCAGGCCCCCGGGCCACACAGGCAAGAGGAGACTCCCGGGCCCTCAGGGCCCTCTCTGC 1346
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
|||||
Db 1347 TTCAGGGGACTTCAGGGCACCGTTGGGGAGCGTGGGGTGGCTGGACCTCGGGGACTGCCA 1406
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
|||||
Db 1407 GGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCGCCCTCTCTGGCCA 1466
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1467 TCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCGCCAGGAGCAAT 1526
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
|||||
Db 1527 AGCTGCCCGCTCCTCCTGGAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAA 1586
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
Db 1587 GAAATTTTGGAGGATGCAAGCTTTTCTGTGAAGACAAGTCTTACATCTTGTTCATA 1646
QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
Db 1647 AACACTAGAGAGAGACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGCCACTGG 1706
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
Db 1707 ATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGGTGGTGGTGGGACATCTCCA 1766
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
Db 1767 GACTACAAAAATGGAAGCTGGACAGCCGATAAAGTGGGTGCTATGGCCATGGCCAGGA 1826
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 1827 GAAGACTGCTGGGTGATTTATGCTGGCAGTGGCAACGATTTCCAATGTGAAGACGTC 1886
QY 441 AsnAspPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
|||||
Db 1887 AATAACTTCATTGGCAAAAGACAGGAGACAGTACTGTCTGCTGATTA 1937

RESULT 3
AB038518 3058 bp mRNA linear PRI 08-MAR-2001
LOCUS
DEFINITION
Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type 1, complete cds.
ACCESSION
AB038518
VERSION
AB038518.1 GI:13365514

KEYWORDS

SOURCE Homo sapiens tissue_lib:Placenta cDNA to mRNA.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites)
Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.

AUTHORS

Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a

TITLE

scavenger receptor family

JOURNAL

Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)

MEDLINE

21092718

REFERENCE

2 (bases 1 to 3058)

Nakamura,K. and Nakamura,T.

Direct Submission

Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical

School, Division of Biochemistry, Biomedical Research Center; 2-2

yamadaoka, Suita, Osaka 565-0871, Japan

(E-mail:knakamura@bich.med.osaka-u.ac.jp,

Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)

Location/Qualifiers

1. .3058

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_lib="Placenta"

1. .3058

/gene="SRCL"

209. .2437

/gene="SRCL"

/codon_start=1

/product="scavenger receptor with C-type lectin type I"

/protein_id="BAB39147.1"

/db_xref="GI:13365515"

/translation="MKDDFAEEEVQSPFYKRFQIGTQCTCKKNWALKPSITILY

ILCALLTITVALIGYKVKEDMNVGTGMETSQTQYDDKLITAVESDLKLGDGTGKAI

STNSLSFTSRDILDLRQLREITEKTSKNDLEKQAGDALVDROSQLEKLENN

SFLITVNTKLTQAGNGYVTNLQDDTSVLQGNLQNGYSHNVVIMNLNLTQVOQR

LITNLSQSDTSQAIQIKNDPQNLQVFOAKKDDTLWLEKQVSLQTLAANNQALA

KANNDTLEDMNSLNSFTQGMENITTSOANEQNLKDLQDLHKDAENRTAIFNOLLE

RPLQETDIVNIIISNYSYTAHLRLTSLNLEVRTCTDTLTKHDDTLTSLNPTLANI

RUDSVLRQDLMLSRDLTEVANLSVIMEKMLVDSKHGOLIKNFTILOGPPGPRGP

RGDRSQPPGPTGNKQKGEKGERGPGPAGERGPIGPAGPPGERGKSGSGPKP

GSGSPGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG

LPGVPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG

KEIFEDAKIFCDKSSHLVFIINTREEQWIKQWVGRESHWIGLDSERENWKLWDG

TSPDYKNNKAGQPDNMGHGHGPGEDCAGLIYAGQWDFQCEDVNNFICEKDETVLSS

AL"

polyA_signal 3020. .3025

BASE COUNT 911 a 744 c 764 g 639 t

ORIGIN

Alignment Scores:

Pred. No.: 4,28e-82 Length: 3058

Score: 2514.00 Matches: 456

Percent Similarity: 99.78% Conservative: 0

Best Local Similarity: 99.78% Mismatches: 1

Query Match: 99.76% Indels: 0

DB: 9 Gaps: 0

US-09-763-712a-2_copy_91_547 (1-457) x AB038518 (1-3058)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20

Db 1064 ATCAACAGCAGCTCAACTCATTACAGGTCAGATGAGCAACATCACCATTCTCTCAA 1123

QY 21 AlaAsnGluGlnAsnLeuLeuAspLeuGlnAspLeuHisAspAlaGluAsnArgThr 40

Db 1124 GCCAACGACGACAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAAACA 1183

QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60

Db 1184 GCATCAAGTTCAACCACTGGAGGAACCTTCCAGCTCTTTGAGACGGATATTGTGAAC 1243

QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 Db 1244 ATCATTAGCAATATCAGTTACACAGCCACCACCTGCGGACGCTGACCAATCTAAAT 1303

QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db 1304 GAAGTCAGGACCACTTGCACAGATACCTTTACCAACACACAGATGATCTGACCTCTTG 1363

QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 Db 1364 AATAATACCTTGCCCAACATCCGTTTGGAATCTGTTCTCTCAGGATGCACAGATTG 1423

QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 Db 1424 ATCAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTTATGGAAGAAATGAAG 1483

QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 Db 1484 CTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAAAGGTCCACCG 1543

QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyGlyAsn 180
 Db 1544 GGGCCAGGGGTCCTCAAGAGGTGACAGAGATCCAGGAGCCCTGCGCCACTGGCAAC 1603

QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProGlyProGlyGly 200
 Db 1604 AAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAGAGAGGC 1663

QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db 1664 CCAATTGGACAGCTGGTCCCGCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCACG 1723

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlyProGlyProGlyProGly 240
 Db 1724 GGCCCCAAGAGCTCCCGTGGTCCCTGGGAGCCCGGCGCTCAGGCGCCCGACGTGGGGAC 1783

QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlyProGlyProGlyProGly 260
 Db 1784 CCAGGCCCCCGGGCCCCACAGGCAAGAGGACTCCCGGCGCTCAGGCGCCCTCGCTGGC 1843

QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 Db 1844 TTCACAGGACTTCACAGGACCGCTGGGAGCGCTGGGGTGCCTGAGACTCGGGGACTGCCA 1903

QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 Db 1904 GGCTTGCTGGGTACAGCATGCCAGGCCCCCAAGGGCCCCCGGCGCTCTCTGGCCCA 1963

QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 Db 1964 TCAGGAGCGTGGTGGCCCTGCGCCCTGCAGATGAGCAACCCCGCAGGAGGACAAT 2023

QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 Db 2024 AGCTGCCCGCTCACCTGGAGAACTTCACAGACAAATGTACTATTATTTTCAGTTGAGAAA 2083

QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 Db 2084 GAAATTTTTCAGATGCCAAGCTTTTCTGTGAAGACAGTCTTCACATCTCTGTTTTCATA 2143

QY 361 AsnThrArgGluGlnGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisIleTrp 380
 Db 2144 AACACTAGAGAGAAACAGCAATGGATAAAAAACAACAGATGGTAGGAGAGAGAGAGAGAG 2203

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db 2204 ATCGGCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGCTGGATGGGACATCTCCA 2263

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 Db 2264 GACTACAAAAATGGAAAGCTGGACAGCGGATAACTGGGGTCATGGCCATGGGCCAGGA 2323

QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 |||||
 Db 2324 GAAGACTGTGCTGGTGTGATTTATCTGCGCAGTGGACGATTTCCAAATGTGAAGACGTC 2383
 |||||
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 2384 AATAACTTCATTTGGAAAAAGACAGGGAGACGACTGTCTCATCTGCATTA 2434
 |||||
 RESULT 4
 LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
 DEFINITION Sequence 27 from Patent WO0208284.
 ACCESSION AX454442
 VERSION AX454442.1 GI:21713845
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 REFERENCE
 AUTHORS Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A.,
 Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
 Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I.
 and Ye,W.
 TITLE Compositions and methods for the diagnosis and treatment of
 disorders involving angiogenesis
 JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
 Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
 (US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,
 Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
 Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
 ; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
 Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
 I. (US)
 FEATURES
 Location/Qualifiers
 1..2005
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 606 a 493 c 491 g 415 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,84e-81 Length: 2005
 Score: 2487.00 Matches: 455
 Percent Similarity: 99.56% Conservative: 0
 Best Local Similarity: 99.56% Mismatches: 2
 Query Match: 98.69% Indels: 1
 DB: 6 Gaps: 0
 US-09-763-712A-2_COPY_91_547 (1-457) x AX454442 (1-2005)
 QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 |||||
 Db 134 ATGAACAGCCACTCACTCATTCAGCTCAGATGGAGAACATCACCACATCTCTCAA 193
 |||||
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 |||||
 Db 194 GCCAACGAGCAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAGAACA 253
 |||||
 QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 |||||
 Db 254 GCCATCAAGTTCACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 313
 |||||
 QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 |||||
 Db 314 ATCATTAGCAATATCAGTTACACAGCCACACCTGCGGACCGCTCACCAGCAATCTAAAT 373
 |||||
 QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 |||||
 Db 374 GAAGTCAGGACCACTGCACATACCTT-ACCAACACACAGATCATCTGACCTCCTTG 432
 |||||
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 |||||

Db 433 AATAATACCTCGCCCAACATCCGTTTGGATCTGTTTCTCTCAGGATGCAACAAGATTGG 492
 |||||
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 |||||
 Db 493 ATGAGGTCGAGGTAGACACTGAAGTAGCAACTATATCAGTATATATGGAAGAAATGAAG 552
 |||||
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 |||||
 Db 553 CTAGTAGACTCCAAGCATGCTCAGCTCAAGAATTTTACAATACTACAAGGTCCACCG 612
 |||||
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrHis 180
 |||||
 Db 613 GGGCCAGGGGTCCAAAGAGGTACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 672
 |||||
 QY 181 LysGlyGlnLysGlyGluProGlyLysGluProGlyProGlyProAlaGlyGluArgGly 200
 |||||
 Db 673 AAGGCACAAAGAGAGAGAGGGGGAGCCCTGGACCACCTGGCCCTGCGGGTGAGAGAGC 732
 |||||
 QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 |||||
 Db 733 CCAATTGGACCACTGCTCCCGCGAGAGCGTGGCGCAAGGATCTTAAGGCTCCCAAG 792
 |||||
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 |||||
 Db 793 GGGCCAAAGGCTCCCGTGGTTCCCTGGGAAGCGCGCCCTCAGGGCCCTCAGTGGGGAC 852
 |||||
 QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProGly 260
 |||||
 Db 853 CCAGGCCCCCGGGCCACCAGCAAGAGGACTCCCGCGCCCTCAGGGCCCTCTCTGGC 912
 |||||
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 |||||
 Db 913 TTCCAGGACTTTCAGGGCACCGTTGGGAGCGCTGGGGTGCCTGGACTCGGGACTGCCA 972
 |||||
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyPro 300
 |||||
 Db 973 GGTCTGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGGCCCTCTCTGGCCCA 1032
 |||||
 QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 |||||
 Db 1033 TCAGGAGCGGTGCTGCCCTGGCCCTGCAGATGACCAACCCCGCCGAGGAGACAAT 1092
 |||||
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTrpPheSerValGluLys 340
 |||||
 Db 1093 AGCTGCCCGCCCTCACTGGAAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1152
 |||||
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 |||||
 Db 1153 GAAATTTTGGAGGTGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATA 1212
 |||||
 QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 |||||
 Db 1213 AACACTAGAGAGAACAGCAATGGATGATAAAAAACAGATGGTAGGGAGAGAGCCACTGS 1272
 |||||
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 |||||
 Db 1273 ATCGGCTCACAGACTCAGAGCGTGAAAATGAATGGAAGTGGCTGTGATGGACATCTCCA 1332
 |||||
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
 |||||
 Db 1333 GACTCAAAAATGGAAAGCTGACAGCGCGGATAACTGGGGTCATGGCCATGGCCGAGGA 1392
 |||||
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 |||||
 Db 1393 GAAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGACGATTTCCAATCTGAAGAGCTC 1452
 |||||
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 1453 AATAACTTCATTTGGAAAAAGACAGGGAGACAGTACTGCTCATCTGCATTA 1503
 |||||
 RESULT 5
 AX490920
 LOCUS AX490920 2005 bp DNA linear PAT 16-AUG-2002

DEFINITION Sequence 27 from Patent WO0200690.

ACCESSION AX490920

VERSION AX490920.1 GI:22323797

KEYWORDS human.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

Baker, K.P., Ferrara, N., Gerber, H., Gerritsen, M.E., Goddard, A., Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J., Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I. and Ye, W.

Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis

Patent: WO 0200690-A 27 03-JAN-2002;

Genentech, Inc. (US)

FEATURES

source

1..2005

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 606 a 493 c 491 g 415 t

ORIGIN

Alignment Scores:

Pred. No.: 2.84e-81 Length: 2005

Score: 2487.00 Matches: 455

Percent Similarity: 99.56% Conservative: 0

Best Local Similarity: 99.56% Mismatches: 2

Query Match: 98.63% Indels: 1

DB: 6 Gaps: 0

US-09-763-712a-2_copy_91_547 (1-457) x AX490920 (1-2005)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20

Db 134 ATGAACAGCCAGCTCAACTCACTTACAGGTGAGTGGAGAAATCACCCTCTCTCAA 193

QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40

Db 194 GCCAACGAGCAGAACCTGAAGACCTGCAGACTTACACAAAGATGCAGAGAATAAACA 253

QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60

Db 254 GCCATCAAGTTCAACCAACTGAGGAGACGCTTCCAGCTCTTGAGACGGATATGTGAAC 313

QY 61 IleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80

Db 314 ATCATTTAGCAATATCATAGTTACAGACCCACCTCGGACCGCTGACCAATCTTAAAT 373

QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100

Db 374 GAAGTCAGGACCATGTCACAGATACCTT-ACCAACACACAGATGATCTGACCTCTTG 432

QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120

Db 433 AATAATACCTCGCCCAACATCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 492

QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140

Db 493 ATGAGTCGAGGTTACACTGAAGTAGCCCACTTATCATGATTATGGAAGAAATGAAG 552

QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160

Db 553 CTAGTAGACTCCAAGCATGTCAGCTCATCAAGATTTTACAATACTACAAGGTCACCG 612

QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsn 180

Db 613 GGCCCCAGGGGTCACAGAGGTGACAGAGGATCCAGGGACCCCTCGCCCAACTGCAAC 672

QY 181 LysGlyGlnLysGlyGluGluProGlyProGlyProGlyProAlaGlyGluArgGly 200

Db 673 AGGGACAGAAAGGAGAGGGGGAGCGCTGGACCACTGGGCCCTCGGGGTGAGAGAGC 732

QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220

Db 733 CCAATTTGGACAGCTGGTCCCTCCCGGAGAGCGTGGCGCAAGGATCTAAAGCTCCAG 792

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240

Db 793 GGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGGAAAGCCCGGCTCAGGGCCCGAGTGGGAC 852

QY 241 ProGlyProProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProProGly 260

Db 853 CCAGGCCCCCGGGGCCACAGGCAAGAGGAGACTCCCGGCCCTCAGGGCCCTCCTGGC 912

QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280

Db 913 TTCCAGGAGACTTCAGGGCACCGTGGGAGACCTGGGGTGCCTGGACCTCGGGACTGCCA 972

QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300

Db 973 GGCTTGCCTGGGTACACAGGCATGCCAGGCCCAAGGGCCCGCCCGGCTCCTGGCCCA 1032

QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320

Db 1033 TCAGGAGCGGTGGTCCCTCGCTGCAGATGAGCCCAACCCCGCACCGGAGGACAAT 1092

QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340

Db 1093 AGCTGCCCGCCCTCAGTGGAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 1152

QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 360

Db 1153 GAAATTTTGGAGATGCAAGGCTTTCTGTGAAGCAAGTCTTCACATCTGTGTTTCATA 1212

QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380

Db 1213 ACACCTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCCACTGG 1272

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400

Db 1273 ATCGGCTCACAGACTCACAGCGTGAAATGAATGGAAGTGGTGGATGGACATCTCCA 1332

QY 401 AspThrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420

Db 1333 GACTACAAAAATTTGAAAGCTGGACAGCCGATAAATGGGGTTCATGGCCATGGCCAGCA 1392

QY 421 GluAspCysAlaGlyLeuIleThrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440

Db 1393 GAAGACTGTCTGGGTGATTATGCTGGCAGTGGAAAGATTTTCCAATGTGAAGACGTC 1452

QY 441 AsnAsnPheIleCysGlyLysAspArgGluThrValLeuSerSerAlaLeu 457

Db 1453 AATAACTTCATTTGGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 1503

RESULT 6

AB078434

LOCUS

DEFINITION Mus musculus CL-P1 mRNA for collectin placenta 1, complete cds.

ACCESSION AB078434

VERSION AB078434.1 GI:21901968

KEYWORDS

SOURCE

ORGANISM

Mus musculus tissue_lib:Liver cdna to mRNA.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1

Ohtani, K., Suzuki, Y., Eba, S., Kawai, T., Kase, T., Keshi, H.,

Sakai, Y., Fukuo, A., Sakamoto, T., Itabe, H., Suzutani, T.,

Ogasawara, M., Yoshida, I., and Wakamiya, N.

cDNA cloning of mouse CL-P1 gene

Unpublished

JOURNAL

REFERENCE

2 (bases 1 to 2637)

Ohtani, K., Suzuki, Y., Eba, S., Kawai, T., Kase, T., Keshi, H.,

Sakai, Y., Fukuo, A., Sakamoto, T. and Wakamiya, N.

TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College, Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp, Tel: 81-166-68-2393, Fax: 81-166-68-2399)
FEATURES Location/Qualifiers
source 1. 2637
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_lib="Liver"
gene 1. 2637
 /gene="CL-P1"
CDS 92..2320
 /gene="CL-P1"
 /codon_start=1
 /product="collectin placenta 1"
 /protein_id="BAC05523.1"
 /db_xref="GI:21901969"
 /translation="MKDDFAEEBEVQFGYKRFQIQGTQCTCKNNWALKFSLVLLY
 ILCALLITVAILGYKVERMDNVTDGMTSHQTYDNKLTAVESDLKLGDQAGKAL
 STNSELSTFRSDLDLRLQQLQEITEKTSKNKDTLEKIQANGDSLVDKQSLKETLQNN
 SFLTITVNTKTLQAVNGYVITLQDQTSVLQNLQSOYTSQSVIMNLTNTVOVOOR
 LINSNQSDVDTSLAIQRIKNDQFONLQOVFLQAKKDDLDLKEKVSQSLQTLAANNALA
 KANNDTLEDNNSLSSETSGQMDNITTIISQANEQSLKDLQDLKDTERTAVKFSQLEE
 RFQVFEDIVNIGNISYTAHLATLISLNDVRICTDILTRHTDLDLTSLNLTSLVNI
 RLDLSLRMQDDMMRSKLDTEVANLSVVMEEMLKLVDSKHGOLIKNFTILQPPGPRGP
 KGRDSQGPFGPTGNKGKGEKGPPEPPGAGERTIGPVPPGERGSKGSKGSGQPK
 GSRGSPGPGQSGPSPGPPGPPGKDLPGPPGPPGQFQGLQGTGVPGPVGPGRGLPG
 LPVPVGMGPKGPPGPPGSGAMEPLALONEPTPASEVNCSPPHKWNFTDKCYVFSLE
 KEIPEDAKLCEDEKSSHLVFINREEQOWIKKHTVGRSHWIGLITSEQSEKWLQDG
 SPVDKNNKAGQPDNWSGSHGPGEDCAGLIYAGQNDPQCDENINFCEREAAPVPS
 IL"
BASE COUNT 829 a 617 c 635 g 556 t
ORIGIN
 Alignment Scores:
 Pred. No.: 4.44e-76 Length: 2637
 Score: 2344.00 Matches: 417
 Percent Similarity: 96.50% Conservativeness: 24
 Best Local Similarity: 91.25% Mismatches: 16
 Query Match: 93.02% Indels: 0
 DB: 10 Gaps: 0
 US-09-763-712a-2_copy_91_547 (1-457) x AB078434 (1-2637)
 Qy 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnLeuThrThrIleSerGln 20
 Db ATGAATACCGAGCTCAGCTCATTCACAGTCAGATGGACATATACCACTATTCACAG 1006
 Qy 21 AlaAsnGluGlnAsnLeuAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 Db GCCAACGACGAGAGCTCAAGACCTTCAGGACCTTACACAGGATACAGAAATAGAACA 1066
 Qy 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 Db GCTGTCAAGATTTCAGCAACTTTCAGGAGCGCTTCCAGGCTTTGAGACAGATATTGTGAC 1126
 Qy 61 IleIleSerAsnIleSerThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 Db ATCATTCACATCAGCTACACAGCCCATCCTTCAGGACACTGACCAAGCAATCTGAAT 1186
 Qy 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db GATGTTAGGACCATGACACAGACCTTGACGACACACAGGATGACCTGACCTCCTTG 1246
 Qy 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 Db AATAACACACTAGTCAACATCCGCTTGGATTCTATTCTCTCAGGATGACGACAGACATG 1306
 Qy 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 Db ATGAGGTCAAAAGTTAGACACTGAAGTGGCCCACTTATCAGTGGTTATGGAACAGATGAA 1366

Qy 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 Db CTGGTTGACTCCAACACGGTCAGCTCATCAAGAACTTTACCATTTCTACAAGGTCTCTCT 1426
 Qy 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlnGlyProProGlyProThrGlyAsn 180
 Db GGCCCCAGAGGTCCAAAGGTGACAGAGATCTCAGGACCACTGGTGGTCCAACTGGCAAC 1486
 Qy 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
 Db AAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
 Qy 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db ACAATTGGACAGTCGCGCCCTCTGGAGAGCGTGGCAGCAAGAGATCCAAAGGCTCACAG 1606
 Qy 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 Db GTCTCCAAAGGATCTGCTGGGTCCCGAGGAAGCTGCGCCCTCAAGGACCTAGTGGGGAC 1666
 Qy 241 ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
 Db CCAGGACCAAGGTCACAGGCAAGGATGACTCCCTGGCCCTCAGGGCCCTCTCTGGC 1726
 Qy 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 Db TTCAGGGGACTACAGGGGACTGTGGGTGAGCTGAGGACTGAGGACTGAGGACTGAGGACT 1786
 Qy 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 Db GCTTGGCAGGGGTGCCAGGCAATGCTGGCTTCAAGGACCACTGGGCTTCCAGGCCCC 1846
 Qy 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 Db TCAGGAGCAATGGAGCCATTCGCTTCAGAGATGAACACACCCAGCATCAGAGGTCAAC 1906
 Qy 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 Db GGATGTCGCGCTCACTGGAAGAACTTCACAGATAAATGCTACTATTTTCATTGGAAAAA 1966
 Qy 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLysLeuValPheIle 360
 Db GAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTTCCCATCTCTCTTTTCATA 2026
 Qy 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 Db AACTCAAGAGAGACACACCATGGATGATAAAGACATACCGTGGGAGAGAAAGCATTTGG 2086
 Qy 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db ATCGCCCTCACAGACTCAGAACAGGAAAGCAATGGAAGTGGCTAGACGGGTCACTGTT 2146
 Qy 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 Db GATTACAAAACTGGAAGCTGGACACACAGATAACTGGGCGAGTGGCCATGGGCACGGA 2206
 Qy 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 Db GAAGACTGTGCTGGCTTGATTTACCGAGACAGTGGAGATGCTTCAGTGTGTGAATC 2266
 Qy 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 Db AATACTTCAITTTGTGAGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2317
RESULT 7
AB038519
LOCUS Mus musculus src1 mRNA for scavenger receptor with C-type lectin, complete cds.
DEFINITION AB038519.1 GI:18146951
ACCESSION AB038519
VERSION AB038519.1
KEYWORDS Mus musculus cDNA to mRNA.
SOURCE

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 AUTHORS Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T.
 TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin
 (SRL)(1), a novel member of the scavenger receptor family
 Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
 21575692

REFERENCE 2 (bases 1 to 3291)
 AUTHORS Nakamura, K. and Nakamura, T.
 TITLE Direct Submission
 JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
 School, Division of Biochemistry, Biomedical Research Center; 2-2
 Yamadaoka, Suita, Osaka 565-0871, Japan
 (E-mail: knakamura@onbich.med.osaka-u.ac.jp,
 Tel.:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)

FEATURES
 source
 1..3291 /organism="Mus musculus"
 /db_xref="taxon:10090"
 gene
 1..3291 /gene="srl"
 CDS
 77..2305 /gene="srl"
 /codon_start=1
 /product="scavenger receptor with C-type lectin"
 /protein_id="BAB82497.1"
 /db_xref="GI:1814692"
 /translation="MKDDFAEEEVQSFYKRFGLHGTCTQCTKINNWLKFSIVLLY
 ILCAITITVAILGYKVKRMNDVSDGNETSHQTYDNKLTAVESDLKLGDAQKKAL
 STNSLSFTFDIDLRQLOEITEKTSKNDLTLEKLAAGDSLVDRQSOLKETQNN
 SFLITTVAKTLQAVNGYVNLQDNTNVLQNSQSVYIMNNLNLTVQOORN
 LITSNQSVDDTSLAIORIKNDFONLQOVFLQAKDTPDLKEKVQSLQTLAANNSALA
 KANNDTLEDMSQLSFTGQMDNLTITISQANEQSLKLDLQDLKHDTENTAYKFSOLEE
 RFQVETDIWINSISNTAHRLFLTSLNNDVWTCTDITLHRTDDTLTSLNLTIVNI
 RLDLSLRKQDMKSLKIDTEVANLSVYMEEMKLVDSKHGQLINKFTILQPPGPRGP
 KGRSQSPGPTGNKGQKEGPEPGPAGERGTIGVPGPGERGSKGSGQPK
 GSRSPGPKQSGPDGPPGPKGDLGPPGPPGPGFQGLQTVGEVGPVGPRLPG
 LPVPGMPKPGPPGPPGSGAMEPLALQNEPTPASEVNGCPPPHWKNFTDKCYFSL
 KEILDAKLFCEKSSHLVFTNSREEQWIKKHTVRESHWIGLTDSPQSEEMKWLGD
 SPVDYKNKKAGQPDNMGSGHGPGECDAGLIYAGQWDPQCDDEINNFICEKEREAVPSS
 IL"

polyA_signal
 3246..3251 /gene="srl"

BASE COUNT 1013 a 757 c 759 g 762 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.66e-75 Length: 3291
 Score: 2330.00 Matches: 415
 Percent Similarity: 96.06% Conservative: 24
 Best Local Similarity: 90.81% Mismatches: 18
 Query Match: 92.46% Indels: 0
 Db: 10 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x AB038519 (1-3291)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 Db 932 ATGAATAGCCAGCTTCATTCACAGGTCCAGATGGAGCAATACCACTATCTCCACAG 991

QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 Db 992 GCCACGACGACGCTGAAACACCTTCAGGACTTACACAGGATACAGAAATAGAACA 1051

QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 Db 1052 GCTGTCAAGTTCCAGCAACTTCAGGAACGCTTCCAGGTCTTTGAGACAGATATTGTGAAC 1111

QY 61 IleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 Db 1112 ATCATTAGCAACATCAGCTACAGCCCATCATCTGAGGACACTGACCAACCAATCTGAAT 1171

QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db 1172 GATGTTTGGACCACTGCACAGACACCTTGACACAGACACACGATGACCTGACCTCTCTG 1231

QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 Db 1232 AATAACACACTAGTCAACATCGCTTGGATTCTTATTTCTCTCAGGATGACGACGACATG 1291

QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 Db 1292 ATGAGTCAAGTTAGACACTGAAGTGGCCACTTATCAGTTGGTTATGGAAGAGATGAAA 1351

QY 141 LeuValAspSerLysHisGlnGlnLeuIleLysAsnPheThrIleLeuGlnGlnProGly 160
 Db 1352 CTGTTGACTCCAAAGCACGCTCAGCTCATCAAGAACTTTACCATCTTACAAGGTCTCTCT 1411

QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
 Db 1412 GCGCCACAGAGTCCAAAAGGTGACAGAGGATCTCAGGAGCACCTGCTGCTCAACTGGCAAC 1471

QY 181 LysGlyGlnLysGlyGluProGlyGluProGlyProGlyProGlyProAlaGluArgGly 200
 Db 1472 AAGGACACAAAGGAGAGGAGGAGCTGCTCCACCTGGCCCTCGGGCTGAGAGGGC 1531

QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db 1532 ACAATTGGACAGTGGCCCTCTCTGGAGAGCGTGGCGAGCAAGGATCCAAAGGCTCACAG 1591

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 Db 1592 GGTCCCAAGAGTCTCGTGGTCCCGAGGAAAGCTGGCCCTCAAGGACCTAGTGGGGAC 1651

QY 241 ProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProGlyProGly 260
 Db 1652 CCAGGACCAAGGTCACCAAGGAGGAGTGGTCCACCTGGCCCTCAGGCGCTCTCTGGC 1711

QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 Db 1712 TTCCAGGGACTACAGGCACTGTGGTGAGCCTGGAGTACCTGGACCTCGGGGTTGCCA 1771

QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 Db 1772 GGTTCGCCAGGGTGCAGGCATGCTGGCCCTAAGGGACCACTGGCCCTCCAGGCCCC 1831

QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 Db 1832 TCAGGAGCAATGGAGCAATGGCTCTGCAGAAATGAACCAACCCACATCAGAGGTCAAC 1891

QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 Db 1892 GGATGTCCGCCCTCACTGGAGAACTTTCACAGATAAATGCTACTATTTTTCATTGGAAAA 1951

QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 Db 1952 GAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTTCCCATCTCGTTTTTCA 2011

QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 Db 2012 AACTCAAGAGAAGACAGCAATGGATAAAGACATACCGTGGGAGAGAAAGCCATTGG 2071

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db 2072 ATCGGCTCACAGACTCAGACAGAGAAAGCGAATGGAAGTGGCTAGACGGGTCACTGT 2131

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 Db 2132 GATTACAAAACCTGGAAGCTGGACCAACACAGATAAAGTGGGAGGAGTGGCCATGGCCAGGA 2191

QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 Db 2192 GAAGACTGTGCTGGTGTGATTTACCGAGGACAGTGGAAATGACTTCAGTGTGATGAATC 2251


```
QY 441 AsnAspHelleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
|||||
Db 2252 AATAACTTATTTGTGAGAGGAAGGAGGAGGAGTACCATCATCATATTA 2302
|||||
RESULT 8
HS0802985 1886 bp mRNA linear PRI 20-MAR-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp547G1215 (from clone DKFZp547G1215).
DEFINITION AL713657
ACCESSION AL713657
VERSION AL713657.1 GI:19584339
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1886)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weill,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp547G1215) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFZp547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbr1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1886
/gene="DKFZp547G1215"
<1..1256
/gene="DKFZp547G1215"
/note="similarity to asialoglycoprotein receptor (Mus
musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKENOLEERFQLPETDIVNIISVYTAHLRLTSLNLEVRT
TCTDTLRKTDLTSLNLTANIRLDSVSLRMOQDLMSRLDTEVANSVIMEKLV
DSKHGLIKNTILQPGPRGDRSGQPPGTGNKGKQKEGEPGPGPAGERG
PIGAPPPGERGGKSGSKGSPKSGRSGFGPGSGSDPGPGPKGELGPGPG
PFGQLGTGEPGVPGRPLGLPGVPMGPKPGPPGPPGSAVPLALONEPTA
PDNGCPHWNFTDKCYFVSEKEIFEDAKLFCDSKSHLVFINTREEQWIKQMV
GRESHWIGLTDSEENENWKLDGTSPDYKNKAGQPDNWHGHGPGEDCAGLIVAGW
NDFOCEDVNNFICEKRDRETVLSSAL"
polya_signal 1838..1843
/gene="DKFZp547G1215"
polya_site 1869
/gene="DKFZp547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Alignment Scores:
Pred. No.: 6.08e-75 Length: 1886
Score: 2309.00 Matches: 416
Percent Similarity: 99.76% Conservative: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 91.63% Indels: 0
DB: 9 Gaps: 0
```

```
US-09-763-712a-2_copy_91_547 (1-457) x HSM802985 (1-1886)
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
|||||
Db 3 GCCATCAAGTTCAACCACTGGAGGAAACGCTTCCAGCTCTTGAGACGATATTGTGAAC 62
|||||
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
|||||
Db 63 ATCATTAGCAATATCAGTTACAGACCCACCACCTCGGACGCTGACCAATCTAAAT 122
|||||
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSerLeu 100
|||||
Db 123 GAAGTCAGACCACTTGCACATACCTTACCAACACACAGATGATGTACCTCTCTTG 182
|||||
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
|||||
Db 183 AATAATACCTCGCCCAACATCGTTTGGATTCTGTTTCTCTCAGGATGCAACAAGATTTC 242
|||||
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
|||||
Db 243 ATGAGTCCAGGTTAGACACTGAAGTAGCAACTTATCAGTGATTATGGAAGAAATGAAG 302
|||||
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
|||||
Db 303 CTAGTAGACTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATACTACAAGGTCCACCG 362
|||||
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsp 180
|||||
Db 363 GGCCTCCAGGGTCCCAAGAGTGACAGAGATCCAGGAGACCCCTGCTGCCCACTGGCAAC 422
|||||
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProGlyGluArgGly 200
|||||
Db 423 AAGGACACAAAGGAGAGAGAAGGGAGCGCTGGACCACTGGCCCTCGGGTGAGAGAGGC 482
|||||
QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
|||||
Db 483 CCAATTGGACCACTGCTGCTCCCGGAGAGCGTGGCGCAAGGATCTAAAGGCTCCCAAG 542
|||||
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlyProGlyProGlyAsp 240
|||||
Db 543 GGCCTCCAAAGGCTCCCGTGTCCCTGGAGAGCGCGCCCTCAGGGCTCCAGTGGGGAC 602
|||||
QY 241 ProGlyProProGlyProProGlyLysGlyLysGlyLeuProGlyProGlyProGlyPro 260
|||||
Db 603 CCAGGCCCCCGGGCCACCAGCAAAAGGAGACTCCCGCGCCCTCAGGGCCCTCTCTGCG 662
|||||
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
|||||
Db 663 TTCCAGGGACTTCAGGGCACCGCTTGGGGAGCGCTGGGGTGCCTCGGAGACTGCCA 722
|||||
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
|||||
Db 723 GCTTGCCTTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGCCCCA 782
|||||
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
|||||
Db 783 TCAGGAGCGGTGCTGCTCCCTGGCCCTGCAGAAATGAGCAACCCCGCCAGCGGAGGACAAT 842
|||||
QY 321 GlyCysProProHisThrLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
|||||
Db 843 GGTGCTCCCGCCCTCACTGGAAGAACTTCACAGACAAATGTACTATTATTTTCAGTTGAGAA 902
|||||
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 360
|||||
Db 903 GAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTGTGTTTCATA 962
|||||
QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
|||||
Db 963 AACACTAGAGAGGAACAGCAATGGATAAAAAACACAGATGGTAGGAGAGAGAGCCACTGG 1022
|||||
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
|||||
Db 1023 ATCGGCTTCACAGACTCAGAGCGTGAAATGAAATGGAAGTGGCTGGATGGGACATCTCCA 1082
|||||
```

```

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
Db 1083 GACTACAAAAATGGAAAGCTGACAGCGGATAAAGTGGGTCTATGCCATGGCCAGGA 1142
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 1143 GAAGACTGTGCTGGGTGATTTATGTGGGCAGTGAACGATTTTCCAAATGTGAAGACGTC 1202
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValIleSerSerAlaLeu 457
Db 1203 AATAACTTATTGGCAAAAAGACAGGGACAGTAGTACTGTCTCTGCATTA 1253

RESULT 9
AB052103
LOCUS Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
DEFINITION type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
REFERENCE
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Keiji Nakamura, Osaka University Graduate
School of Medicine, Division of Biochemistry, Biomedical Research
Center, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)
FEATURES
source
1..4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..4330
/gene="SRCL"
3..1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:13365553"
/translation="MKDDFAEEVEVPSFGKRFGEHGTQCPKHNNAKFSIILY
ILCALLTITVAIGLYKVEKWDNVGGMETSQTYDDKLITAVESDLKLIGDQTKKAI
STNSLSTERSDILDRQRLREITEKSKNDLLEKQASGDALVDROSQKLETLENN
SFLITVNKTLOAGNYVTNLQDDTSVLOGNQMYSHNVVIMNINLNTQVOQRN
LITNQRVDVTSQAIQIKNDQFNLOQVFLQAKKDTDLKKEVQSLOTLAANNSALA
KANNDLEDMSQNSFTGQMNITISQANEONKLDQLHDKDAENRTAIKNOLEE
RPFQETDIVNISVTAHLRTLNLNEVRTTCTDILT KHTDDLTSLNNTLANI
RLDSVLSMOODLRSRLDTEVANLSVIMEKMLVDSKHGQLIKNFTILOQPPGRGP
RGDRSQSPGPTGNKQKGEKGPPEPAGERPIGPAGPBERGKSGKSGGPK
GSRGSPGKPGQSGDPGPGKGLPGPPGPGFQGLQGTGEPGVPGRGLPG
LPGVPMGPGRGPPGPGPSGAVPLALQNEPTPAEDNSKSPSLQPGGGSACA"
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN
Alignment Scores:
Pred. No.: 2,41e-72 Length: 4330
Score: 2244.00 Matches: 457
Percent Similarity: 46.21% Conservative: 0
Best Local Similarity: 46.21% Mismatches: 0
Query Match: 89.05% Indels: 532
DB: 9 Gaps: 1

```

```

US-09-763-712a-2_COPY_91_547 (1-457) x AB052103 (1-4330)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetClnAsnIleThrThrIleSerGln 20
Db 858 ATGAACAGCCAGCTCAACTCATTCACAGGTACAGTACAGAACATCACCATATCTCTCAA 917
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 918 GCCAACAGCAGACAACTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAGAACA 977
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 978 GCCATCAAGTTCAACCAACTGGAGGAGCGCTTCCAGCTCTTTGAGACGGATATGTGAAC 1037
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
Db 1038 ATCATTAGCAATATCAGTTACACAGCCACCCACTCGGAGCTGACCAATCTAAAT 1097
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 1098 GAAATCAGGACCACTTGCACAGATACCTTACCAAAACACACAGATGATCTGACCTCTTG 1157
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 1158 AATAATACCTTGCCCAACATCCGTTTGGATTCTTCTCTCAGGATGCAACAAGATTG 1217
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
Db 1218 ATGAGGTCAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGNAGAAATGAAG 1277
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
Db 1278 CTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGATTTTACAATACTACAAGGTCCACCG 1337
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 180
Db 1338 GGCCCCAGGGGTCCAGAGGTGACAGAGGATCCAGGGACCCCTGCCCCCACTGGCAAC 1397
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
Db 1398 AAGGACACAAAGGAGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAGAGAGGC 1457
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
Db 1458 CCAATTGGACCACTGGTCCCTCCCGGAGAGCGTGGCGGCAAAAGGATCTAAAGCTCCAG 1517
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
Db 1518 GGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAGAGCCCGGCGCTCAGGGCCCCCAGTGGGAC 1577
QY 241 ProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProGly 260
Db 1578 CAGAGCCCCCGGGGCCACACAGGCAAGAGGAGTCCCGGGCCCTCAGGGCCCTCTCTGGC 1637
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1638 TTCCAGGGACTTCAGGGCACCGTGGGGAGCGCTGGGTGCCTGGACCTCGGGGACTGCCA 1697
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
Db 1698 GGCTTTCCTTGGGTACCAAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCTCTCTGCCCA 1757
QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1758 TCAGGACCGGTGGTGGCCCTGCGCTGAGAGATGACCAACCCCGGACCGGAGGACAAT 1817
QY 320 ----- 320
Db 1818 AGTAAGTCCAAGCCCTCCCTCCAGCCAGGGGACAGGGCAGTGCATGTGCTTGAGACTGA 1877
QY 320 ----- 320

```

Db 1878 GCCAACCCCACTGTGTGGCCCCCAACCATGTTATTTTGTGTTTGTGTTTGTGAGAC 1937
QY 320 ----- 320
Db 1938 AGGGTCTCACCTGTGTGCCAGGCTGGAGTCAGTGGCGTGATCTCGGCTCACTGCAACC 1997
QY 320 ----- 320
Db 1998 CCCACATCCGGGTTCAGTGATCTCCTGCTTAGCCTCCAGTAGCTGGGAATACAA 2057
QY 320 ----- 320
Db 2058 GCGTGTGCCCCACAGCCAGCTAATTTTGTATTTTATAGAGAGGGGTTTCACAGTGT 2117
QY 320 ----- 320
Db 2118 TGGCAGGGTGTTCACAACTCTGGCCTCAAGTCACCTACCTAGTTTCCCAAGT 2177
QY 320 ----- 320
Db 2178 GCTGGGAATACAGGTGTGAACCAACACCGTTTGTGCCCGGTGCATATTTAATTAACAGC 2237
QY 320 ----- 320
Db 2238 AAAGCATCATAGGGGTTGGCACAGGAGCCCTGTGTCAGAACAGGCCCTCCATACCAAGC 2297
QY 320 ----- 320
Db 2298 ATGTGCTCTCTCAACGCCATTTGGGTTTTTGTAGTGCATTTTGTATGTTATATGT 2357
QY 320 ----- 320
Db 2358 GTGTATGTATGGACAGGGTCTCACTCTGTCAACAGGCTGGAATGACGAAGTAG 2417
QY 320 ----- 320
Db 2418 TGCAACATAGCTCATTCACAGTTTGACCTCTGGGCTTAGTGATCTTCCACCTCAGAC 2477
QY 320 ----- 320
Db 2478 TCCTCGTAGCTGGACCAAGGGTATGCCACACACCTGGCTAAATTTTTTATTTTCT 2537
QY 320 ----- 320
Db 2538 TTTTGTAGAAACAGGGTCTTGCTATGTGCCCCAGGCTGTGAACTCTCGGCTCAAG 2597
QY 320 ----- 320
Db 2598 CAACCTCCTGCCTCAGTCTTGCRAAGTGTGGGATTCAGGTGGGAGCCACTGCACCTG 2657
QY 320 ----- 320
Db 2658 GCTCTTGCACTGTGTTATTTATAGAGCCAGCCCTCAGAAGACAAACACATGCAAAACA 2717
QY 320 ----- 320
Db 2718 TTTCTTCAGGAAGCCTGCTGGCCGCTGAGCGCTCTCCCTCGTTTATACCTCCAC 2777
QY 320 ----- 320
Db 2778 TCTGGAGATAGTGGCTCCCTCGGTAGTCCATCAACATGTGGATGGAACCCCTCTGCT 2837
QY 320 ----- 320
Db 2838 ATTGATGACTACATTAATTTCTTTTGGAGAAAGTGTTAGATGTTACGGTGCCTTCTAG 2897
QY 320 ----- 320
Db 2898 TCCTTTGATAATTAACCTTTTGCAAACTCGAATGTTTGTGTAGCAACCAAGCCTATT 2957
QY 320 ----- 320
Db 2958 TTCCTCTCTTGACCAACCCGCTATCTACTGTGATTTTGGTGGTGGAGCCGGGTATT 3017

QY 320 ----- 320
Db 3018 TTTGTTGATCCGTGAGCCCATTTTGTGCTGGGTTCTCACATCCAAGTGTCTCTGGGTT 3077
QY 320 ----- 320
Db 3078 CCGTATGATTCAGACTGACGCTGAACATCTGTGCTGCCCTGCTTCTTGTGTTGCTTC 3137
QY 320 ----- 320
Db 3138 TGTCAATTTCTGTCATTTGTGAAGACGGATCCAACACAGAAGAACAAAGACCCAGTCCAAG 3197
QY 320 ----- 320
Db 3198 ATCCATCCAGATTATGTGGTTTTCGGCTTGTGCTCTGACAGTCCACTTTGGGTGCTATGCT 3257
QY 320 ----- 320
Db 3258 TGGAGCAGCTCCGTGAACGTCAGCTGTGGGCTCGTGCAGGAGCGCTCAGGCTGCTCCC 3317
QY 320 ----- 320
Db 3318 AGCGGACCAAGTATTGGCCCTCTGAACACACAGCTGAAGAAAGGACAGAGTGTCTCTGTG 3377
QY 321 -----
Db 3378 ATCTTTTAATCAATGTTTTCACACTTTTGCATTCCTAGGCTGCCGCCCTCAGTGAAGAAC 3437
QY 329 PheThrAspLysCysTyrPheSerValGluLysGluIlePheGluAspAlaLysLeu 348
Db 3438 TTCACAGACAAATGCTACTATTTTTCAGTTGAGAAGAAATTTTGTAGGATGCAAGCTT 3497
QY 349 PheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnP 368
Db 3498 TTCGTGAAGCAAGTCTTCACATCTTGTTCATAAACACTAGAGAGAACAGCAATGG 3557
QY 369 IleLysLysGlnMetValGlyArgGluSerHisThrPheGlyLeuThrAspSerGluArg 388
Db 3558 ATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGT 3617
QY 389 GluAsnGluThrLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGly 408
Db 3618 GAAATGAATGAAAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGA 3677
QY 409 GlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyr 428
Db 3678 CAGCCGGATAACTGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTAT 3737
QY 429 AlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAsp 448
Db 3738 GCTGGCAGTGAACGATTTCCAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAAGAC 3797
QY 449 ArgGluThrValLeuSerSerAlaLeu 457
Db 3798 AGGGAGACAGTACTGTCTATGTCATTA 3824

RESULT 10

AC112416

LOCUS

DEFINITION

AC112416

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Mammalia;

Eutheria;

Sciurognathi;

Muridae;

Rattus.

1 (bases 1 to 130763)

REFERENCE

AUTHORS

AC112416 130763 bp DNA linear HTG 13-JUL-2002
Rattus norvegicus clone CH230-350K4, *** SEQUENCING IN PROGRESS
***, 28 unordered pieces.

AC112416

AC112416.3 GI:21743403

HTG; HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa;

Chordata; Vertebrata;

Euteleostomi;

Mammalia;

Eutheria;

Sciurognathi;

Muridae;

Rattus.

1 (bases 1 to 130763)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Ali-Osman,F.R., Allen,C.,

Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Eathwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, K., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louiseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Wei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Soneike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vexa, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 130763)
 Worley, K.C.

Direct Submission
 Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 130763)
 Worley, K.C.

Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:20303233.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GRRG
 Center clone name: CH230-350K4
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 113686 bases at least Q40
 Consensus quality: 115729 bases at least Q30
 Consensus quality: 117044 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1103: contig of 1103 bp in length
 * 1104 1203: gap of unknown length
 * 1204 3068: contig of 1865 bp in length
 * 3069 3168: gap of unknown length
 * 3169 4755: contig of 1587 bp in length
 * 4756 4855: gap of unknown length
 * 4856 7046: contig of 2191 bp in length
 * 7047 7146: gap of unknown length
 * 7147 8651: contig of 1505 bp in length
 * 8652 8751: gap of unknown length
 * 8752 10001: contig of 1250 bp in length
 * 10002 10101: gap of unknown length
 * 10102 11278: contig of 1177 bp in length
 * 11279 11378: gap of unknown length
 * 11379 13651: contig of 2273 bp in length
 * 13652 13751: gap of unknown length
 * 13752 15942: contig of 2191 bp in length
 * 15943 16042: gap of unknown length
 * 16043 17619: contig of 1577 bp in length
 * 17620 20301: contig of unknown length
 * 20302 20401: gap of unknown length
 * 20402 23598: contig of 3197 bp in length
 * 23599 23698: gap of unknown length
 * 23699 26125: contig of 2427 bp in length
 * 26126 26225: gap of unknown length
 * 26226 30263: contig of 4038 bp in length
 * 30264 32964: gap of unknown length
 * 32965 33064: contig of 2601 bp in length
 * 33065 36455: gap of unknown length
 * 36456 40431: contig of 3876 bp in length
 * 40432 40531: gap of unknown length
 * 40532 42635: contig of 2104 bp in length
 * 42636 42735: gap of unknown length
 * 42736 46230: contig of 3495 bp in length
 * 46231 46331: gap of unknown length
 * 46331 49593: contig of 3263 bp in length
 * 49594 49694: gap of unknown length
 * 49694 55037: contig of 5343 bp in length
 * 55037 55137: gap of unknown length
 * 55137 60936: contig of 5800 bp in length
 * 60937 61037: gap of unknown length
 * 61037 69030: contig of 7994 bp in length
 * 69031 69130: gap of unknown length
 * 69131 79184: contig of 10054 bp in length
 * 79185 79284: gap of unknown length
 * 79285 91807: contig of 12523 bp in length
 * 91808 91907: gap of unknown length
 * 91908 106422: contig of 14515 bp in length
 * 106423 106522: gap of unknown length
 * 106523 119310: contig of 12788 bp in length
 * 119311 119410: gap of unknown length
 * 119411 130763: contig of 11353 bp in length.

FEATURES

Location/Qualifiers
 1. 130763
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-350K4"

BASE COUNT 38110 a 25956 c 25888 g 38088 t 2721 others
 ORIGIN

Alignment Scores:

Pred. No.: 1.39e-27 Length: 130763
 Score: 1021.50 Matches: 215
 Percent Similarity: 41.47% Conservative: 16

Best Local Similarity:	38.60%	Mismatches:	37
Query Match:	40.54%	Indels:	289
DB:	2	Gaps:	5

US-09-763-712A-2_COPY_91_547 (1-457) x AC112416 (1-130763)

Qy	153	PheThrIleLeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGln	172
Db	13787	TTTTCTTTATGTTTAAAGTCTCTCTGGGCCACAGAGGTCCAAAAGGTGACAGAGGATCTCAG	13846
Qy	173	GlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProProGlyPro	192
Db	13847	GGACCGCTGGTCCAACTGGCCACACAGGGACAGAAAGGAGAGAGAGACCCCTGGTCCA	13906
Qy	193	ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGly	212
Db	13907	CCTGGCCCTGGGGTGAGCGAGGCACAAATTTGGACCAGTAGGCCCTCTCTGGAGAGCGTGGC	13966
Qy	213	GlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro	232
Db	13967	AGCAAAAGATCCAAAGGCTCACAGGGTCCCAAGAGATCACGTGGGTCCCCAGGAAAGGCT	14026
Qy	233	GlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeu	252
Db	14027	GGCCCTCAAGGGCCTAGTGGAGACCTTGGACCACCAAGGTCCACCAGGCAAGGATGGACTC	14086
Qy	253	ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly	272
Db	14087	CCTGGCCCTCAGGGCCCTCTCTGGCTTCACGGGACTACAGGCACTGTGGGTGAACCTGGA	14146
Qy	273	ValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys	292
Db	14147	GTACTTGGACCTCGGGGGCTACAGAGCTTACAGGGGTGCCAGGCATGCCCTGGGCCCAAG	14206
Qy	293	GlyProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGlu	312
Db	14207	GGACCACTGGCCCCCAGAGTCCCTCAGGAGCAATGGAGCCGTTGGCTCTGCAGAAATCAA	14266
Qy	313	ProThrProAlaProGluAspAsn-	320
Db	14267	CCAAACCCAGCATCAGAGGTCAACGGTAAGTCCAAATCATCCTTTTCAGCCAGAAATAAG	14326
Qy	320	-----	320
Db	14327	GCAGTGTGTAGGCTTGGCTAAACCAGACCTTCAGGAGTCCCTTAATTACACTTAATATCAG	14386
Qy	320	-----	320
Db	14387	GGACCCCTATGTGTTCTCTTGGATGCCCTGGTGTCTCATCTCAGGACCCCTGTGTTCTCTGCT	14446
Qy	320	-----	320
Db	14447	TAAACATTATCAGACTCTACTCATTTAGATGAGCGAGGCCCTCAGAAAGATACCTATACC	14506
Qy	320	-----	320
Db	14507	CCAAGTACTTATTTCAAGCTATCTGCACCGGTGCTCCCTGACCATCTCTTCTTGCTCT	14566
Qy	320	-----	320
Db	14567	TCCTTCTTCTCTAGAAACAGACCATTTCCATCAATTTAACTCCATCAGTATGCAGATGGG	14626
Qy	320	-----	320
Db	14627	TTTCTGTGATAGACACACTCTGGTTTGTGCTTAAGAAAGCTGCTGGGTGCTCTGCTTTTT	14686
Qy	320	-----	320
Db	14687	GACCATTAAAGCTTTCACAAAACCTCTGTCTAGTTAGGAAACCAAGCCATGTGGTCACTCA	14746
Qy	320	-----	320
Db	14747	TGCTGGTCCATGAACCTTTGGTTCTGAAGTACACTGTTTACAAGCTCCAGCTGATCCAAAG	14806

QY	320	-----	320	
Db	14807	ACTGACAAATGAGTATCAGGGCCACCCCACTTCTCTTCTGTGTCATTTATAAGCCTAAC	14866	
QY	320	-----	320	
Db	14867	CAACCCCAATCACAGCCCTTCCACACTGTGTGGTCTTCACTATTGTGTACTGTAAATCC	14926	
QY	320	-----	320	
Db	14927	ACAACGTGTGAACATCTGCCATGTCTCTACAGGGACACTCCATGGCATGCAAGCAGTG	14986	
QY	320	-----	320	
Db	14987	TGAGGCTCTACTTTGGTATCTTAAACATAGCTGAGGAAATGCAGAGTGCCTTTTTTTTT	15046	
QY	321	-----GlyCysProHisTrpLysAsnPhethr	330	
Db	15047	AGATCAACATTTAAAAACGTTACATTCCTAGGATGTCGCCCTCACTGGAAGACTTCACA	15106	
QY	331	AspLysCysTrpTyrPheSerValGluLysGluLlePheGluAspAlaLysLeuPheCys	350	
Db	15107	GATAAATGTTACTATTTTTTCGTCGAGAAAGAAATTTTTGAAGATGCTAAAGCTTTCTGT	15166	
QY	351	GluAspLysSerHisLeuValPheIleAsnThrArgGluGluGln-----	366	
Db	15167	GAAGACAATCTCCCATCTCGTTTTTCATAAATCAAGAGAGAACAGGTATGCAGCTG	15226	
QY	367	---GlnTrpIleLysLysGlnMetValGlyArg-----GluSerHisTrpIleGly	382	
Db	15227	CATCTTGGTTGGAGAAAACTATAATTTAGAGATAAGTTATACTTTGAAAAATGCATGCCT	15286	
QY	383	LeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyr	402	
Db	15287	TTGAGGGAGTCAAGAGAGATCAAAAATTCATGATCTCATAAACAGGAAGCTGCAGTTG	15346	
QY	403	LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAsp	422	
Db	15347	-----CATGGCTTTATCCAGGCATGCGA	15370	
QY	423	CysAlaGlyLeu-----IleTyrAlaGlyGlnTrp	432	
Db	15371	ATATGTGAATGTCTATGCTTCTTAGCGCCACAGATATATAGTCCACCTGG	15421	
RESULT 11	AC114677	182029 bp	DNA linear	HTG 10-JUN-2002
LOCUS	Mus musculus clone RP24-213K19, WORKING DRAFT SEQUENCE, 10 ordered pieces.			
DEFINITION	AC114677			
ACCESSION	AC114677			
VERSION	AC114677.3	GI:21362159		
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 182029)			
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.			
JOURNAL	Mus musculus, clone RP24-213K19			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 182029)			
	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kellis, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,			

*	22425	22524:	gap of	100	bp
*	22525	38094:	contig of	15370	bp
*	38095	38194:	gap of	100	bp
*	38195	49220:	contig of	11026	bp
*	49221	49320:	gap of	100	bp
*	49321	66394:	contig of	17074	bp
*	66395	66494:	gap of	100	bp
*	66495	91692:	contig of	25198	bp
*	91693	91792:	gap of	100	bp
*	91793	120869:	contig of	29077	bp
*	120870	120969:	gap of	100	bp
*	120970	169088:	contig of	48119	bp

FEATURES

```

location/quadri11ers
1. .169088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-324C2"
/clone_lib="RPC1-11 Human Male BAC"
1. .17943
/note="assembly_fragment
clone_end:SP6
vector_side:left"
18044.22424
/note="assembly_fragment"
22525.38094
/note="assembly_fragment"
38195.49220
/note="assembly_fragment"
49321.66394
/note="assembly_fragment"
66495.91692
/note="assembly_fragment"
91793.120869
/note="assembly_fragment"
120970.169088
/note="assembly_fragment
clone_end:r7
vector_side:right"
48103 a 33537 c 34088 g 52660 t 700 others
BASE COUNT
ORIGIN

```

Alignment Scores:

Pred. No.:	4.16e-25	Length:	169088
Score:	954.50	Matches:	191
Percent Similarity:	58.40%	Conservative:	21
Best Local Similarity:	52.62%	Mismatches:	53
Query Match:	37.88%	Indels:	98
DB:	2	Gaps:	8

US-09-763-712A-2_COPY_91_547 (1-457) x AC016128 (1-169088)

Qy	131	AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGly-----Gln	148
		: : : : : : : :	
Db	153903	AAPTTCGAGATTATTATPAAGAGAAAAATATTATCAATCAATCATTCACAGTGGATTTTCAT	153844
Qy	149	LeuIleLysAsnPhethrIleLeuGlnGlyProGlyProArgGlyProArgGlyAsp	168
		: : :	
Db	153843	GTTGACACCTTTTCTTTTATGCTTTAGTCTCACCGGGCCCCAGGGGTCACAGAGGTGAC	153784
Qy	169	ArgGlySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlnLysGly	188
Db	153783	AGAGGATCCAGGGAGCCCTCGGCCACACTGSCACACAGGGACACAAAGAGGAGAGAGGGG	153724
Qy	189	GluProGlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro	208
Db	153723	GRGCTGGACCACTTGGCCCTCGGGGTGAGAGGCCCAATTGGACCACTGGTCCCCC	153664
Qy	209	GlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySer	228
Db	153663	GGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCCAAGGGCCCCAAAGGCTCCCGTGGTTC	153604

Qy	229	ProGlyLysProGlnGlyProSerGlyAspProGlyProGlyProGlyProGly	248
Ds	153603	CCTGGGAAGCCCGGCCCTCAGGGCTCCAGTGGGACCCAGGCCCGGCCACCAGGC	153544
Qy	249	LysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrVal	268
Ds	153543	AAAGAGGACTCCCGGCCCTCAGGGCCCTCCCTGGGTTCAGGGACTTCAGGGCACCGTT	153484
Qy	269	GlyGluProGlyValProGlyProAArgGlyLeuProGlyLeuProGlyValProGlyMet	288
Ds	153483	GGGAGGCTGGGTGCTTGACCTCGGGGACTGCCAGGCTTGCTGGGTACCAAGCATG	153424
Qy	289	ProGlyProLysGlyProProGlyProProGlyProSetGlyAlaValProLeuAla	308
Ds	153423	CCAGGCCCAAGGCCCGGCCCTCCCTGGCCATCAGGAGCGTGGTGGCCCTGGCC	153364
Qy	309	LeuGlnAsnGluProThrProAlaProGluAspAsnGly-----	321
Ds	153363	CTGCAGAATGAGCCAACCCAGCACCGGAGACAATGGTAAGTCCAAGCCCTCCCTCCAG	153304
Qy	322	-----CysPro--ProHil 325 :	
Ds	153303	CCAGGGGACAGGCAGTCATGTCTTGAGACTGAGCCNACCCCCAGTGTGTGGCCCCA	153244
Qy	325	sTrpLysAsnPheThrAspLysCysTyrtYrPheSerValGluLysGluIlePheGluAs	345
Ds	153243	C-----CCATGTTATTTT----- 153231 :	
Qy	345	pAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGl	365
Ds	153230	-TGTTTTTGTGTTTTGTGTTTCAGACAGGGTCTCACTCTGTT----- 153192 :	
Qy	365	uGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHistripIleGlyLeuThrAs	385
Ds	153192	----- 153192	
Qy	385	pSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTryLysAsnTr	405
Ds	153191	-GCCAGGCTGGAGTCAGTGGCGGTGATCTCGGCTCAGTCAACCTCCACATCCCGGGTT	153133
Qy	405	pLys-----AlaglyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGl	421
Ds	153132	CAAGTGATTCTCCTGCCTTAGCCTCCAGGACTAGCTGGGACTACAAGCGTGTGCCCCACCA	153073
Qy	421	uAspCysAlaGlyLeuIleTyrrAlaGlyGln-----TrpAs 433 :	
Ds	153072	GCCAGCTAAATTTTGTATTTTGTAGTAGAGAGGGGGTTTCAACGCTGTGCCAGGCTGGTC	153013
Qy	433	n-----AspPheGl 436 : :	
Ds	153012	TCAAACCTCTGGCCTCAAGTGATCCACCTACCTTAGTCTCCCAAAGTGTGGGATTACAG	152953
Qy	436	nCysGlu 438 :	
Ds	152952	GTGTGAG 152946	
RESULT 14			
AP000915/c			
LOCUS	AP000915	Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete sequences.	PRI 27-APR-20
DEFINITION			
ACCESSION	AP000915.5	GI:20334314	
VERSION	AP000915		
KEYWORDS	HTG.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Homo sapiens genomic DNA		
JOURNAL	Published Only in Database (1999)		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		

REFERENCE	AUTHORS	TITLE	JOURNAL
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

QY	325	strPlysAsnPhetrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAs	345
Db	44364	C-----CCATGTTATTT-----	44352
QY	345	pAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluG1	365
Db	44351	-TGTGTTTGTGTTTGTGTTTGTGACAGGCTCTCACTCTGTT-----	44313
QY	365	uGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAs	385
Db	44313	-----	44313
QY	385	pSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTr	405
Db	44312	-GCCAGCTGGAGTGCAGTGGCGGTGATCTCGGCTCACTGCAACCTCCACATCCCGGGTT	44254
QY	405	pLys-----AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyG1	421
Db	44253	CAAGTGATTCCTCGCTTACGCTCCAGGTAGCTGGGACTACAAGCGTGTGCCCCCA	44194
QY	421	uAspCysAlaGlyLeuIleTrpAlaGlyGln-----	433
Db	44193	GCCAGCTAATTTTGTGTTATTTTAGTAGAGAGGGGTTTCACCGTGTGGCAGCTCGTC	44134
QY	433	n-----	436
Db	44133	TCAAACCTCTGGGCTCAAGTGATCCACCTACCTTAGTCTCCAAAGTGCTGGGATTACAG	44074
QY	436	nCysGlu	438
Db	44073	GTGTGG	44067
RESULT	15		
ACCESSION	AC024368		
LOCUS	Homo sapiens chromosome 11 clone RP11-179K3 map 11, LOW-PASS		
DEFINITION	SEQUENCE SAMPLING.		
VERSION	AC024368.1	GI:7108157	
KEYWORDS	HTG; HTGS_PHASE0.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 71044)		
JOURNAL	Blair, B., Linton, L., Nusbaum, C. and Lander, E.		
REFERENCE	Homo sapiens chromosome 11, clone RP11-179K3		
AUTHORS	2 (bases 1 to 71044)		
	Blair, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,		
	Anderson, S., Baldwin, J., Barna, N., Bada, F., Boguslavsky, L.,		
	Bouknight, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,		
	Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,		
	DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M.,		
	Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,		
	Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,		
	Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,		
	Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,		
	Klein, J., Landers, T., Largoocke, K., Lehoczy, J., Levine, R.,		
	Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,		
	McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,		
	Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Naylor, J.,		
	Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,		
	Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,		
	Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,		
	Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,		
	Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,		
	Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,		
	Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and		
	Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome		

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted February 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6145
 Center clone name: 179_K_3

* NOTE: This record contains 90 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 699 798: contig of 698 bp in length
 799 798: gap of 100 bp
 1495: contig of 697 bp in length
 1496 1595: gap of 100 bp
 1596 2284: contig of 689 bp in length
 2285 2384: gap of 100 bp
 2385 3056: contig of 672 bp in length
 3057 3156: gap of 100 bp
 3157 3851: contig of 695 bp in length
 3852 3951: gap of 100 bp
 3952 4627: contig of 676 bp in length
 4628 4727: gap of 100 bp
 4728 5409: contig of 682 bp in length
 5410 5509: gap of 100 bp
 5510 6183: contig of 674 bp in length
 6184 6283: gap of 100 bp
 6284 6984: contig of 701 bp in length
 6985 7084: gap of 100 bp
 7085 7768: contig of 684 bp in length
 7769 7868: gap of 100 bp
 7869 8550: contig of 682 bp in length
 8551 8650: gap of 100 bp
 8651 9352: contig of 702 bp in length
 9353 9452: gap of 100 bp
 9453 10140: contig of 688 bp in length
 10141 10240: gap of 100 bp
 10241 10922: contig of 682 bp in length
 10923 11022: gap of 100 bp
 11023 11709: contig of 687 bp in length
 11710 11809: gap of 100 bp
 11810 12511: contig of 702 bp in length
 12512 12611: gap of 100 bp
 12612 13304: contig of 693 bp in length
 13305 13404: gap of 100 bp
 13405 14095: contig of 691 bp in length
 14096 14195: gap of 100 bp
 14196 14881: contig of 686 bp in length
 14882 14981: gap of 100 bp
 14982 15674: contig of 693 bp in length
 15675 15774: gap of 100 bp
 15775 16460: contig of 686 bp in length
 16461 16560: gap of 100 bp
 16561 17250: contig of 690 bp in length
 17251 17350: gap of 100 bp
 17351 18040: contig of 690 bp in length
 18041 18140: gap of 100 bp
 18141 18836: contig of 696 bp in length
 18837 18936: gap of 100 bp
 18937 19621: contig of 685 bp in length

* 19622 19721: gap of 100 bp
 19722 20425: contig of 704 bp in length
 20426 20525: gap of 100 bp
 20526 21224: contig of 699 bp in length
 21225 21324: gap of 100 bp
 21325 22017: contig of 693 bp in length
 22018 22117: gap of 100 bp
 22118 22815: contig of 698 bp in length
 22816 22915: gap of 100 bp
 22916 23609: contig of 694 bp in length
 23610 23709: gap of 100 bp
 23710 24402: contig of 693 bp in length
 24403 24502: gap of 100 bp
 24503 25191: contig of 689 bp in length
 25192 25291: gap of 100 bp
 25292 25981: contig of 690 bp in length
 25982 26081: gap of 100 bp
 26082 26774: contig of 693 bp in length
 26775 26874: gap of 100 bp
 26875 27563: contig of 689 bp in length
 27564 27663: gap of 100 bp
 27664 28355: contig of 692 bp in length
 28356 28455: gap of 100 bp
 28456 29145: contig of 690 bp in length
 29146 29245: gap of 100 bp
 29246 29943: contig of 698 bp in length
 29944 30043: gap of 100 bp
 30044 30745: contig of 702 bp in length
 30746 30845: gap of 100 bp
 30846 31535: contig of 690 bp in length
 31536 31635: gap of 100 bp
 31636 32329: contig of 694 bp in length
 32330 32429: gap of 100 bp
 32430 33114: contig of 685 bp in length
 33115 33214: gap of 100 bp
 33215 33898: contig of 684 bp in length
 33899 33998: gap of 100 bp
 33999 34694: contig of 696 bp in length
 34695 34794: gap of 100 bp
 34795 35486: contig of 692 bp in length
 35487 35586: gap of 100 bp
 35587 36292: contig of 706 bp in length
 36293 36392: gap of 100 bp
 36393 37095: contig of 703 bp in length
 37096 37195: gap of 100 bp
 37196 37872: contig of 677 bp in length
 37873 37972: gap of 100 bp
 37973 38663: contig of 691 bp in length
 38664 38763: gap of 100 bp
 38764 39449: contig of 686 bp in length
 39450 39549: gap of 100 bp
 39550 40236: contig of 687 bp in length
 40237 40336: gap of 100 bp
 40337 41025: contig of 689 bp in length
 41026 41125: gap of 100 bp
 41126 41818: contig of 693 bp in length
 41819 41918: gap of 100 bp
 41919 42608: contig of 690 bp in length
 42609 42708: gap of 100 bp
 42709 43399: contig of 691 bp in length
 43400 43499: gap of 100 bp
 43500 44169: contig of 670 bp in length
 44170 44269: gap of 100 bp
 44270 44966: contig of 697 bp in length
 44967 45066: gap of 100 bp
 45067 45750: contig of 684 bp in length
 45751 45850: gap of 100 bp
 45851 46545: contig of 695 bp in length
 46546 46645: gap of 100 bp
 46646 47336: contig of 691 bp in length
 47337 47436: gap of 100 bp
 47437 48123: contig of 687 bp in length
 48124 48223: gap of 100 bp

Alignment Scores:

US-09-763-712A-2_COPY_91_547 (1-457) x AC024368 (1-71044)

Search completed: March 21, 2003, 11:17:28
Job time : 3321.43 secs

